**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| (51) International Patent Classification 5 : C12N 15/00, C12P 19/34, 21/06 C07H 21/00 | A1 | (11) International Publication Number: WO 91/16427 |
|---|---|---|
| | | (43) International Publication Date: 31 October 1991 (31.10.91) |

(21) International Application Number: PCT/US91/02910

(22) International Filing Date: 24 April 1991 (24.04.91)

(30) Priority data:
513,957 24 April 1990 (24.04.90) US

(71) Applicant: STRATAGENE [US/US]; 11099 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: SHORT, Jay, M. ; 320 Delage Drive, Encinitas, CA 92024 (US). SORGE, Joseph, A. ; 17021 Circa Del Sur, Rancho Santa Fe, CA 92067 (US).

(74) Agents: BIGGS, Suzanne, L. et al.; 611 West Sixth Street, 34th Floor, Los Angeles, CA 90017 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, SE (European patent).

**Published**
*With international search report.*

(54) Title: METHODS FOR PHENOTYPE CREATION FROM MULTIPLE GENE POPULATIONS

(57) Abstract

Methods of producing biological agents which express a desired identifiable phenotype are provided. These methods include bringing together populations of diverse replicas of nucleotide sequences to give a plurality of combined nucleotide sequences, each comprising one member of each population, expressing the combined nucleotide sequences to give a phenotype and identifying those biological agents expressing the desired phenotype.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| **AT** | Austria | **ES** | Spain | **MG** | Madagascar |
| **AU** | Australia | **FI** | Finland | **ML** | Mali |
| **BB** | Barbados | **FR** | France | **MN** | Mongolia |
| **BE** | Belgium | **GA** | Gabon | **MR** | Mauritania |
| **BF** | Burkina Faso | **GB** | United Kingdom | **MW** | Malawi |
| **BG** | Bulgaria | **GN** | Guinea | **NL** | Netherlands |
| **BJ** | Benin | **GR** | Greece | **NO** | Norway |
| **BR** | Brazil | **HU** | Hungary | **PL** | Poland |
| **CA** | Canada | **IT** | Italy | **RO** | Romania |
| **CF** | Central African Republic | **JP** | Japan | **SD** | Sudan |
| **CG** | Congo | **KP** | Democratic People's Republic of Korea | **SE** | Sweden |
| **CH** | Switzerland | | | **SN** | Senegal |
| **CI** | Côte d'Ivoire | **KR** | Republic of Korea | **SU** | Soviet Union |
| **CM** | Cameroon | **LI** | Liechtenstein | **TD** | Chad |
| **CS** | Czechoslovakia | **LK** | Sri Lanka | **TG** | Togo |
| **DE** | Germany | **LU** | Luxembourg | **US** | United States of America |
| **DK** | Denmark | **MC** | Monaco | | |

# DESCRIPTION

## Methods for Phenotype Creation From Multiple Gene Populations

### Cross Reference to Related Application

This is a continuation-in-part application of copending application Serial No. 513,957, filed April 24, 1990 which is a continuation-in-part of Serial No. 353,235, filed May 16, 1989, and Serial No. 353,241, filed May 17, 1989, the disclosures of which are hereby incorporated by reference.

### Field of the Invention

The present invention relates to methods for randomly combining populations of nucleotide sequences and selecting those combinations coding for a desired predetermined phenotype.

### Background of the Invention

The production of genetic variants, including variants of both polypeptides and organisms such as bacteria and phage, has been a goal in the work of many individuals involved in recombinant DNA technologies. For example, researchers have beneficially relied upon random genetic recombination in the past for the production of new and useful microorganisms. Genetic recombination includes a variety of processes that produce new linkage relationships of genes or parts of genes. Genetic recombination is often subdivided into general genetic recombination, which takes place between homologous chromosomes, more or less anywhere along their length, and recombination that does not require extensive homology. The latter category includes site-specific recombination, which depends upon the existence of specific sites in one or more molecules and which includes interactions of viral genomes and insertion sequences with chromosomes of prokaryotes and

eukaryotes, and less well defined instances of recombination that appear to require neither extensive homology nor special sites. Variable gene expression can also result in production of various combinations of polypeptides, the immune system being one example of such protein combination.

The immune system of a mammal is one of the most versatile biological systems; probably greater than 1.0 x $10^7$ antibody specificities can be produced. Indeed, a great deal of contemporary biological and medical research is directed toward tapping this repertoire. During the last decade, furthermore, there has been a dramatic increase in the ability to harness the output of the immune system. The development of the hybridoma methodology by Kohler and Milstein has made it possible to produce monoclonal antibodies, i.e., a composition of antibody molecules of single epitope specificity, from the repertoire of antibodies induced during an immune response. Monoclonal antibodies have been generated in the past from hybridomas, generated by fusing antibody-secreting lymphocytes with an immortal cell line, such as myeloma.

Although standard hybridoma technology has been extremely valuable, the screening of fused cells to identify hybridomas expressing useful antibody molecules is labor intensive, time consuming and expensive. Moreover, the standard technology yields rodent antibody molecules that have two clear disadvantages. The first is that subtle variations in certain human antigenic systems, such as major histocompatibility proteins, are not easily distinguished by non-primate antibodies. Therefore, rodent antibodies may not provide the repertoire of specificities needed to distinguish certain polymorphic antigenic determinants. In other words, current methods for generating monoclonal antibodies are not capable of efficiently surveying the entire antibody response induced by a particular immunogen. Thus, in an individual animal there

are at least 5-10,000 different B-cell clones capable of generating unique antibodies to a small relatively rigid immunogens, such as, for example dinitrophenol. Further, because of the process of somatic mutation during the generation of antibody diversity, essentially an unlimited number of unique antibody molecules may be generated. In contrast to this vast potential for different antibodies, current hybridoma methodologies typically yield only a few hundred different monoclonal antibodies per fusion. A second major drawback in hybridoma technology is that rodent antibodies are highly immunogenic in humans, and can preclude their continued use in patients for diagnostic or therapeutic purposes.

One alternative is to produce human cells that express antibody. Unfortunately, it is quite difficult to identify and produce pure human monoclonal antibodies. Standard methods used to immortalize antibody-producing cells are less than satisfactory. One approach that circumvents the need for human hybridoma cells has been to use recombinant DNA technology to express fusion antibody proteins. These molecules have amino terminal variable domains of the light and heavy chains derived from a specific rodent monoclonal antibody and the carboxy terminal constant region domains derived from a human antibody. The use of human constant regions diminishes the human anti-globulin immune response, avoiding the stimulation of anti-isotypic antibody-producing B cells. However, the rodent-derived variable region framework domains still elicit a response that is more severe than a variable domain response directed against a pure human antibody.

In an effort to avoid the anti-idiotypic response directed against the rodent framework regions of the domains, some researchers have taken a human antibody and replaced the hypervariable regions (CDRs) with hypervariable regions from a rodent antibody specific for a selected antigen. Although such antibodies may have an

affinity for antigen comparable to the parent rodent antibody, the process of grafting all rodent CDRs into a human immunoglobulin gene is technically challenging.

Aside from repertoire specificity and immunogenicity, other drawbacks in producing monoclonal antibodies with the hybridoma methodology include genetic instability and low production capacity of hybridoma cultures. One means by which the art has attempted to overcome these latter two problems has been to clone the immunoglobulin-producing genes from a particular hybridoma of interest into a procaryotic expression system. See, for example, Robinson et al., PCT Publication No. WO 89/0099; Winter et al., European Patent Publication No. 0239400; Reading, U.S. Patent No. 4,714,681; and Cabilly et al., European Patent Publication No. 0125023.

The immunologic repertoire of vertebrates has recently been found to contain genes coding for immunoglobulins having catalytic activity. Tramontano et al., Sci., 234:1566-1570 (1986); Pollack et al., Sci., 234:1570-1573 (1986); Janda et al., Sci., 244:437-440 (1989). The presence of, or the ability to induce the repertoire to produce, antibody molecules capable of a catalyzing chemical reaction, i.e., acting like enzymes, had previously been postulated almost 20 years ago by W. P. Jencks in Catalysis in Chemistry and Enzymology, McGraw-Hill, N.Y. (1969).

It is believed that one reason the art failed to isolate catalytic antibodies from the immunological repertoire earlier, and its failure to isolate many to date even after their actual discovery, is the inability to screen a large portion of the repertoire for the desired activity. Another reason is believed to be the bias of currently available screening techniques, such as the hybridoma technique, towards the production high affinity antibodies inherently designed for participation in the process of neutralization, as opposed to catalysis.

In an attempt to enhance the designed recombination of desired DNA sequences or the desired combination of otherwise randomly generated polypeptides, including the identification and production of pure human monoclonal antibodies, we have pursued alternative approaches for the production and screening of such nucleotide sequences and polypeptides.

Summary of the Invention

The present invention is directed to methods for producing biological agents having a desired novel phenotype wherein this phenotype results from expression of a particular combined nucleotide sequence and wherein said phenotype can be used to identify the biological agents having the particular combined nucleotide sequence and distinguish them from biological agents having other combined nucleotide sequences. The desired phenotype is typically a phenotype which is not normally expressed by the parent nucleotide sequences. In one embodiments these methods comprise first replicating at least portions of two parent nucleotide sequences. The replicating step yields a population of diverse replicas of parent nucleotide sequences. In one embodiment, each parent nucleotide sequence initially comprises a population (or family) of diverse nucleotide sequences which is replicated to give a population of diverse replicas. Alternatively, a population of diverse replicas is generated by replicating a parent nucleotide sequence under conditions which allow mutations to occur which generates diversity from one parent nucleotide sequence and results in a population of diverse replicas. In one aspect, the parent nucleotide sequences may comprise a single DNA molecule or alternatively the parent nucleotide sequences comprise separate DNA molecules. Where the parent nucleotide sequences comprise one DNA molecule, after replication, the resulting populations of diverse replicas derived from each parent nucleotide sequence are separated. The populations of

diverse replicas are then brought together, preferably in a random manner, to produce combined nucleotide sequences wherein each combined nucleotide sequence comprises one member of each population of diverse replicas. The parent nucleotide sequences may be suitably replicated and brought together according to the various methods described herein for replication and recombination of nucleotide sequences and generation of combinatorial libraries. The combined nucleotide sequences are expressed in biological agents. Such biological agents may comprise a host cell, or alternatively, a plasmid, bacteriophage or virus, or nucleic acid vector, and such suitable means for expression are described herein. In one embodiment< expression may constitute the mere existence of th enucleotide sequences in the same biological agent. Then, the biological agents which express the desired phenotype are identified. If desired, the phenotype is used to distinguish those biological agents expressing the particular combined nucleotide sequence from biological agents expressing other combined nucleotide sequences. The desired phenotype may comprise a polypeptide, more than one polypeptide, or a multimeric polypeptide, the expression of which is detectable. Alternatively, the phenotype may comprise synthesis of one or more RNA molecules. Optionally, either the polypeptides or RNA may exhibit enzymatic activity or receptor activity; or the DNA or RNA may simply act as a target for interaction with other molecules.

The present invention provides novel methods for the cloning of cells having novel phenotypes. These methods generally include the use of a combinatorial library selection system to generate a diverse collection of clones. In one aspect, the methods utilize at least two starting populations of nucleotide sequences which can be recombined to form a library of clones containing nucleotide sequences from each of the parent populations. These methods can be utilized, therefore, to create cells having

novel phenotypes, that is, cells having a new and desired combination of expressed polypeptides. These methods can also be used for the production of new combinations of polypeptides, including the polypeptides utilized in the formation of biologically competent immunoglobulin molecules. In accordance with the latter object of the invention, these methods can be used to screen a larger portion of the immunological repertoire for receptors having a preselected activity than has heretofore been possible, thereby overcoming the before-mentioned inadequacies of the hybridoma technique.

In another embodiment, the present invention contemplates a gene library comprising an isolated admixture of at least about $10^3$, preferably at least about $10^4$ and more preferably at least $10^5$ $V_H$-and/or $V_L$-coding DNA homologs, a plurality of which share a conserved antigenic determinant. Preferably, the homologs are present in a medium suitable for <u>in vitro</u> manipulation, such as water, phosphate buffered saline and the like, which maintains the biological activity of the homologs.

In one embodiment, at least two starting populations of DNA sequence-containing vectors are physically combined by any of several techniques, including those described herein, to form a library of clones containing DNA sequences from each of the parent populations. Alternatively there may be more than two gene families and the vectors produced thereby may contain a random assortment of one member of each gene family to create the identifiable characteristic. These vectors can then be transferred to desired host cells to create <u>in vivo</u> novel combinations of phenotypic characteristics in the host cell. Methods of combining desired DNA sequences include the use of restriction digestion and ligation, homologous recombination, and site-specific recombination by methods including intergrase-related proteins, flp recombinase-catalyzed recombination, the <u>cre</u>-<u>lox</u> system of bacteriophage P1, and the use of transposons.

In a still further embodiment, the present invention contemplates vectors for use in the methods which comprise, in addition to random DNA sequences from the starting gene family populations, DNA sequences which facilitate the region-specific, random recombination together of at least one gene from each starting gene family population. Sequences enabling the recombination of these vectors include the use of functional flp recombination sequences, functional loxp recombination sequences, at sequences recognized by integrase-related proteins from lambdoid bacteriophages, and terminal repeat sequences recognized by transposases. Thus, the present invention also includes methods for the combinatorial generation of phenotypes, including a method of producing a nucleic acid vector encoding two or more desired genes each from a family of genes, said genes being capable of producing a characteristic that can be used to identify the vector encoding said genes from other vectors encoding other members of the families of genes, which method comprises:

a) randomly inserting into vectors one member of a first family of genes and one member from one or more other families of genes so that a population of vectors are created wherein each vector may contain one of the genes from said first gene family and one of the genes from each of said other gene families;

b) identifying within said population of vectors a vector capable of detectably producing a desired characteristic resulting from the inclusion of one gene from said first gene family and one gene from each of said other gene families, and using said characteristic to distinguish the vector from other vectors within the population containing undesired combinations of gene members from said gene families.

Suitable vectors for use according to the methods of the present invention include plasmid or cosmid vectors or, alternatively, phage vectors. Suitable host cells for

expressing the vectors comprise either eukaryotic cells or prokaryotic cells. Preferred eukaryotic cells include mammalian cells. In one preferred aspect, the vectors comprise lambda bacteriophage and host cells comprise E. Coli.

Preferably, the genes are combined in vivo.

Various suitable methods may be used for the identification of a particular vector within the recombinant vector population. These methods include (a) the interaction of sequence-specific nucleic acids with genes from the individual families which were combined: (b) the hybridization of nucleic acid probes with genes from the gene families; (c) the expression of one or both genes from the gene families as an RNA molecule; and (d) the expression of one or both genes as an identifiable protein molecule. Optionally, such an identifiable protein molecule may contain a binding site for another molecule, an epitope recognized by an antibody, or an immune molecule binding site for an epitope. In a preferred identification method, both genes express an RNA and/or polypeptide and said RNAs and/or polypeptides physically interact with a host to create an identifiable characteristic. Both genes may express polypeptides that physically interact to form a neo-epitope recognized by an immune molecule or polypeptides that physically interact to form a binding site for another molecule. Optionally those polypeptides are derived from antibody genes such that the interaction of both polypeptides forms an antigen binding site.

In another preferred aspect, the vectors produced according to the present invention contain a single promoter that expresses the genes from the gene families. Alternatively, the genes from the gene families are each expressed from their own promoter.

In a still further embodiment, the present invention contemplates the creation of combinations of two or more nucleotide sequence families (or populations) by in vitro

recombination. Such in vitro recombination could be carried out using specific recombination target sequences and specific recombinases (like flp recombinase), or by using homologous sequences shared by both nucleotide sequence populations to facilitate homologous recombination.

One method to accomplish a form of homologous recombination in vitro is by using in vitro nucleic acid amplification methods such as the polymerase chain reaction (PCR). If both of two populations of DNA sequences share a region of homology, then it is possible during the PCR for base-pairing to occur between single stranded nucleic acid molecules from both populations of nucleotide sequences. If such base pairing creates a "primer-template complex" that can be used by a polymerase to begin synthesis of complementary strands, then a fusion product is created which will contain sequences from both nucleotide sequence populations (See Figure 21 here). If the shared region of homology is present on most or all of the two nucleotide sequence populations, then most or all of the nucleotide sequences can participate in such recombination. Thus, a combinatorial population of fusion nucleotide sequences can be produced, and subsequently inserted into a single expression vector for expression of the nucleotide sequence from both sequence families. Such a combinatorial population of expressed sequences can then be screened for new phenotypes that would not be present if the sequences from only one population of nucleotide sequences were expressed, and would be present only with expression of particular combinations comprising a nucleotide sequence from each population. For example, such phenotypes could comprise the creation of heterodimeric proteins where one subunit of the dimer is encoded by one nucleotide sequence family and the other subunit of the dimer is encoded by the other nucleotide sequence family.

Thus, the present invention is directed to methods of creating diversity, namely populations of diverse replicas

of nucleotide sequences which may be combined to give a diversity of phenotypes, from which a desired phenotype may be selected. Such diversity may be generated starting with a single DNA molecule which is treated to create diversity, such as by mutagenesis or by starting with a family of nucleotide sequences (or genes) or a combinatorial library.

For example, one may start with a plasmid containing antibody sequences coding for both a light chain and a heavy chain which has been isolated from a known monoclonal-antibody producing cell line. The nucleotide sequences coding for the light chain and the heavy chain may be individually amplified (using a method such as PCR) under conditions that mutated sequences are generated to create a population of mutated sequences. The individual populations of mutated sequences may be used to make combinatorial libraries which are then used to create novel phenotypes. Alternatively, these individual populations of mutated sequences may be combined using techniques such as fusion polynucleotide amplification (for example) fusion PCR (as described herein) and used to generate novel phenotypes. These novel phenotypes may include antibodies having enhanced antigen binding characteristics.

According to another aspect of the present invention, one or more genetically distinct phage may be lytically replicated, conditions which are somewhat mutagenic, to generate a population(s) of diverse phage. Phage having phenotypes distinct from the originals may be generated by cleavage such as by a restriction endonuclease, followed by mixing of phage populations, and ligation, followed by selection for expression of desired phenotypes. In this way phage having diverse phenotypes distinct from the parental phage may be generated combinatorially.

In another embodiment, the methods are utilized to produce novel human antibody-expressing DNA sequences. First, an immunoglobulin heavy chain variable region $V_H$

gene library containing a substantial portion of the $V_H$ gene repertoire of a vertebrate is synthesized. In preferred embodiments, the $V_H$-coding gene library contains at least about $10^3$ and more preferably at least about $10^4$ and more preferably at least about $10^5$ different $V_H$-coding nucleic acid strands referred to herein as $V_H$-coding DNA homologs.

The gene library can be synthesized by various methods, depending on the starting material. Where the starting material is a plurality of $V_H$-coding genes, the repertoire is subjected to two distinct primer extension reactions. The first primer extension reaction uses a first polynucleotide synthesis primer capable of initiating the first reaction by hybridizing to a nucleotide sequence conserved (shared by a plurality of genes) within the repertoire. The first primer extension reaction produces a plurality of different $V_H$-coding homolog complements (nucleic acid strands complementary to the genes in the repertoire). The second primer extension reaction produces, using the complements as templates, a plurality of different $V_H$-coding DNA homologs. The second primer extension reaction uses a second polynucleotide synthesis primer that is capable to initiating the second reaction by hybridizing to a nucleotide sequence conserved among a plurality of $V_H$-coding gene complements.

Where the starting material is a plurality of complements of different $V_H$-coding genes provided by a method other than the first primer extension reaction, the repertoire is subjected to the above-discussed second primer extension reaction. That is, where the starting material is a plurality of different $V_H$-coding gene complements produced by a method such as denaturation of double strand genomic DNA, chemical synthesis and the like, the complements are subjected to a primer extension reaction using a polynucleotide synthesis primer that hybridizes to a plurality of the different $V_H$-coding gene complements provided. Of course, if both a repertoire of

V-coding genes and their complements are present in the starting material, both approaches can be used in combination.

A $V_H$-coding DNA homolog, i.e., a gene coding for a receptor capable of binding the preselected ligand, is then segregated from the library to produce the isolated gene. This may be accomplished by operatively linking for expression a plurality of the different $V_H$-coding DNA homologs of the library to an expression vector. The $V_H$-expression vectors so produced are introduced into a population of compatible host cells, i.e., cells capable to expressing a gene operatively linked for expression to the vector. The transformants are cultured under conditions for expressing the receptor coded for by the $V_H$-coding DNA homolog. The transformants are cloned and the clones are screened for expression of a receptor that binds the preselected ligand. Any of the suitable methods well known in the art for detecting the binding of a ligand to a receptor can be used. A transformant expressing the desired activity is then segregated from the population to produce the isolated gene.

A receptor having a preselected activity produced by a method of the present invention, preferably a $V_H$ or $F_v$ as described herein, is also contemplated.

The present invention also encompasses products produced by the methods of the invention, such as the biological agents produced thereby, also the expression products of these methods such as polypeptides and nucleic acids, vectors produced and kits comprising any of the products of the claimed methods.

Brief Description of the Drawings

In the drawings forming a portion of this disclosure:

Figure 1 illustrates a schematic diagram of the immunoglobulin molecule showing the principal structural features. The circled area on the heavy chain represents the variable region ($V_H$), a polypeptide containing a

biologically active (ligand binding) portion of that region, and a gene coding for that polypeptide, are produced by the methods of the present invention. Sequences L03, L35, L47 and L48 could not be classified into any predefined subgroups.

Figure 2A is a diagrammatic sketch of an H chain of human IgG (IgG1 subclass). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the presence of four domains, each containing an intrachain disulfide bond (S-S) spanning approximately 60 amino acid residues. The symbol CHO stands for carbohydrate. The V region of the heavy (H) chain ($V_H$) resembles $V_L$ in having three hypervariable CDR (not shown).

Figure 2B is a diagrammatic sketch of a human K chain (Panel 1). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the intrachain disulfide bond (S-S) spanning about the same number of amino acid residues in the $V_L$ and $C_L$ domains. Panel 2 shows the locations of the complementarily-determining regions (CDR) in the $V_L$ domain. Segments outside the CDR are the framework segments (FR).

Figure 3 depicts the amino acid sequence of the $V_H$ regions of 19 mouse monoclonal antibodies with specificity for phosphorylcholine. The designation HP indicates that the protein is the product of a hybridoma. The remainder are myeloma proteins. (From Gearhart et al., Nature, 291:29, 1981.)

Figure 4 illustrates the results obtained from PCR amplification of mRNA obtained from the spleen of a mouse immunized with FITC. Lanes R17-R24 correspond to amplification reactions with the unique 5' primers (2-9, Table 1) and the 3' primer (12, Table 1), R16 represents the PCR reaction with the 5' primer containing inosine (10, Table 1) and 3' primer (12, Table 1). Z and R9 are the amplification controls; control Z involves the amplification of $V_H$ for a plasmid (PLR2) and R9 represents the amplification

from the constant regions of spleen mRNA using primers 11 and 13 (Table 1).

Figure 5 depicts nucleotide sequences of clones from the cDNA library of the PCR amplified $V_H$ regions in Lambda ZAP vector. The N-terminal 110 bases are listed here and the underlined nucleotides represent CDR1 (complementary determining region).

Figures 6A and 6B depict the sequence of the synthetic DNA insert inserted into Lambda ZAP vector to produce Lambda Zap II $V_H$ (6A) and Lambda Zap $V_L$ (6B) expression vectors. The various features required for this vector to express the $V_H$ and $V_L$-coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., J. Biol. Chem., 255:27, 1980, and various restriction enzyme sites used to operatively link the $V_H$ and $V_L$ homologs to the expression vector. The $V_H$ expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions heavy chain ($V_H$ Backbone). This $V_H$ Backbone is just upstream and in the proper reading as the $V_H$ DNA homologs that are operatively linked into the Xho I and Spe I. The $V_L$ DNA homologs are operatively linked into the $V_L$ sequence (6B) at the Nco I and Spe I restriction enzyme sites and thus the $V_H$ Backbone region is deleted when the $V_L$ DNA homologs are operatively linked into the $V_L$ vector.

Figure 7 depicts the major features of the bacterial expression vector Lambda Zap II $V_H$ ($V_H$-expression vector) are shown. The synthetic DNA sequence from Figure 6 is shown at the top along with the $T_3$ polymerase promoter from Lambda Zap II vector. The orientation of the insert in Lambda Zap II vector is shown. The $V_H$ DNA homologs are inserted into the Xho I and Spe I restriction enzyme sites. The $V_H$ DNA are inserted into the Xho I and Spe I site and the read through transcription produces the

decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 8 depicts the major features of the bacterial expression vector Lambda Zap II $V_L$ ($V_L$ expression vector) are shown. The synthetic sequence shown in Figure 6B is shown at the top along with the $T_3$ polymerase promoter from Lambda Zap II vector. The orientation of the insert in Lambda Zap vector II is shown. The $V_L$ DNA homologs are inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The $V_L$ DNA homologs are inserted into the Nco I and Spe I cloning sites of the Phagemid.

Figure 9 depicts a modified bacterial expression vector Lambda Zap II $V_L$II. This vector is constructed by inserting this synthetic DNA sequence,

TGAATTCTAAACTAGTCGCCAAGGAGACAGTCATAATGAA

TCGAACTTAAGATTTGATCAGCGGTTCCTCTGTCAGTATTACTT

ATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCTG

TATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGAC

CCCAACCAGCCATGGCCGAGCTCGTCAGTTCTAGAGTTAAGCGGCCG

GGGTTGGTCGGTACCGGCTCGAGCAGTCAAGATCTCAATTCGCCGGCAGCT

into Lambda Zap II vector that has been digested with the restriction enzymes Sac I and Xho I. This sequence contains the Shine-Dalgarno sequence (ribosome binding site), the leader sequence to direct the expressed protein to the periplasm and the appropriate nucleic acid sequence to allow the $V_L$ DNA homologs to be operatively linked into the SacI and XbaI restriction enzyme sites provided by this vector.

Figure 10 depicts the sequence of the synthetic DNA segment inserted into Lambda Zap II vector to produce the lambda $V_L$II-expression vector. The various features and restriction endonuclease recognition sites are shown.

Figure 11 depicts the vectors for expressing $V_H$ and $V_L$ separately and in combination. The various essential components of these vectors are shown. The light chain vector or $V_L$ expression vector can be combined with the $V_H$

expression vector to produce a combinatorial vector containing both $V_H$ and $V_L$ operatively linked for expression to the same promoter.

Figure 12 depicts the labelled proteins immunoprecipitated from E. coli containing a $V_H$ and a $V_L$ DNA homolog are shown. In lane 1, the background proteins immunoprecipitated from E. coli that do not contain a $V_H$ or $V_L$ DNA homolog are shown. Lane 2 contains the $V_H$ protein immunoprecipitated from E. coli containing only a $V_H$ DNA homolog. In lanes 3 and 4, the commigration of a $V_H$ protein a $V_L$ protein immunoprecipitated from E. coli containing both a $V_H$ and a $V_L$ DNA homolog is shown. In lane 5 the presence of $V_H$ protein and $V_L$ protein expressed from the $V_H$ and $V_L$ DNA homologs is demonstrated by the two distinguishable protein species. Lane 5 contains the background proteins immunoprecipitated by anti-E. coli antibodies present in mouse ascites fluid.

Figure 13 depicts the transition state analogue (formula 1) which induces antibodies for hydrolyzing carboxamide substrate (formula 2). The compound of formula 1 containing a glutaryl spacer and a N-hydroxysuccinimide-linker appendage is the form used to couple the hapten (formula 1) to protein carriers KLH and BSA, while the compound of formula 3 is the inhibitor. The phosphonamidate functionality is a mimic of the stereoelectronic features of the transition state for hydrolysis of the amide bond.

Figure 14 illustrates the PCR amplification of Fd and kappa regions from the spleen mRNA of a mouse immunized with NPN. Amplification was performed as described in Example 17 using RNA cDNA hybrids obtained by the reverse transcription of the mRNA with primer specific for amplification of light chain sequences (Table 2) or heavy chain sequences (Table 1). Lanes F1-F8 represent the product of heavy chain amplification reactions with one of each of the eight 5' primers (primers 2-9, Table 1) and the unique 3' primer (primer 15, Table 2). Light chain (k) amplifi-

cations with the 5' primers (primers 3-6, and 12, respectively, Table 2) are shown in lanes F9-F13. A band of 700 bps is seen in all lanes indicating the successful amplification of Fd and k regions.

Figure 15 depicts the screening of phage libraries for antigen binding is depicted according to Example 17C. Duplicate plaque lifts of Fab (filters A,B), heavy chain (filters E,F) and light chain (filters G,H) expression libraries were screened against $^{125}$I-labelled BSA conjugated with NPN at a density of approximately 30,000 plaques per plate. Filters C and D illustrate the duplicate secondary screening of a cored positive from a primary filter A (arrows) as discussed in the text.

Screening employed standard plaque lift methods. XL1 Blue cells infected with phage were incubated on 150mm plates for 4 hours at 37°C, protein expression induced by overlay with nitrocellulose filters soaked in 10mM isopropyl thiogalactoside (IPTG) and the plates incubated at 25° for 8 hours. Duplicate filters were obtained during a second incubation employing the same conditions. Filters were then blocked in a solution of 1% BSA in PBS for 1 hour before incubation with rocking at 25° for 1 hour with a solution of $^{125}$I-labelled BSA conjugated to NPN ($2 \times 10^6$ cpm $ml^{-1}$; BSA concentration at 01 M; approximately 15 NPN per BSA molecule) in 1% BSA/PBS. Background was reduced by pre-centrifugation of stock radiolabelled BSA solution at 100,000 g for 15 minutes and pre-incubation of solutions with plaque lifts from plates containing bacteria infected with a phage having no insert. After labeling, filters were washed repeatedly with PBS/0.05% Tween 20 before development of autoradiographs overnight.

Figure 16 depicts the specificity of antigen binding as shown by competitive inhibition is illustrated according to Example 17C. Filter lifts from positive plaques were exposed to $^{125}$I-BSA-NPN in the presence of increasing concentrations of the inhibitor NPN.

In this study a number of phages correlated with NPN binding as in Figure 15 were spotted (about 100 particles per spot) directly onto a bacterial lawns. The plate was then overlaid with an IPTG-soaked filter and incubated for 19 hours at 25°. The filter were then blocked in 1% BSA in PBS prior to incubation in $^{125}$I-BSA-NPN as described previously in Figure 15 except with the inclusion of varying amounts of NPN in the labeling solution. Other conditions and procedures were as in Figure 15. The results for a phage of moderate affinity are shown in duplicate in the figure. Similar results were obtained for four other phages with some differences in the effective inhibitor concentration ranges.

Figure 17 depicts the characterization of an antigen binding protein is illustrated according to Example 17D. The concentrated partially purified bacterial supernate of an NPN-binding clone was separated by gel filtration and aliquots from each fraction applied to microtitre plates coated with BSA-NPN. Addition of either anti-decapeptide (---) or anti-kappa chain antibodies conjugated with alkaline phosphatase was followed by color development. The arrow indicates the position of elution of a known Fab fragment. The results show that antigen binding is a property of 50 kD protein containing both heavy and light chains.

Single plaques of two-NPN-positive clones (Figure 15) were picked and the plasmid containing the heavy and light chain inserts excised. 500 ml cultures in L-broth were inoculated with 3 ml of a saturated culture containing the excised plasmids and incubated for 4 hours at 37°C. Proteins synthesis was induced by the addition of IPTG to a final concentration of 1mM and the cultures incubated for 10 hours at 25°C. 200 ml of cells supernate were concentrated to 2 ml and applied to a TSK-G4000 column. 50 µl aliquots from the eluted fractions were assayed by ELISA.

For ELISA analysis, microtitre plates were coated with BSA-NPN at 1 ug/ml, 50 $\mu$l samples mixed with 50 $\mu$l PBS-Tween 20 (0.05%)-BSA (0.1%) added and the plates incubated for 2 hours at 25°. After washing with PBS-Tween 20-BSA, 50 $\mu$l of appropriate concentrations of a rabbit anti-decapeptide antibody (20) and a goat anti-mouse kappa light chain (Southern Biotech) antibody conjugated with alkaline phosphatase were added and incubated for 2 hours at 25°. After further washing, 50 $\mu$l of p-nitrophenyl phosphate (1mg/ml in 0.1M Tris pH 9.5 containing 50 mM $MgCl_2$) were added and the plates incubated for 15-30 minutes before reading the OD at 405nm.

Figure 18A depicts the major features of the bacterial expression vector HCFLP containing a $V_H$ DNA homolog and a flp recombination site.

Figure 18B depicts the major features of the bacterial expression vector LCFLP containing a $V_L$ DNA homolog and a flp recombination site properly oriented for recombination with the HCFLP vector.

Figure 19 depicts a diagrammatic sketch of bacterial coinfection with HCFLP and LCFLP vectors for the production of recombinant expression vectors containing $V_L$ and $V_H$ DNA homologs.

Figure 20 depicts an outline showing arm selection for heavy and light chain recombinant vector products using flp recombinase in conjunction with selection based on the inclusion of genes having amber mutations.

Figure 21 shows an outline of a method of phenotype creation using the fusion PCR process described herein.

Figure 22 illustrates human fusion PCR inside primers. The heavy chain $C_H1'$ inside primer sequence is written 3' to 5' and the light chain $V_L$ inside primer sequence is written 5' to 3'. Note that it is not the primer strands that cross-prime to create the fusion molecule, but the complementary PCR product strands. Boxed nucleotides represent regions where the $C_H1'$ primer hybridizes to the 3' end of $C_H1$ on human IgG heavy chain

mRNA or where the $V_L$ primer hybridizes to the 5' end of $V_L$ framework-1 on human kappa light chain cDNA. Underlined sequences indicate the two stop condons. The italicized amino acid and nucleotides indicate changes in sequence from the original pelB leader sequence. The mouse fusion-PCR internal primers overlap in a similar manner.

Figure 23 illustrates an ethidium bromide stained agarose gel. After PCR amplification from human cloned DNA of heavy chain alone (HC), light chain alone (LC), and the heavy/light dicistronic DNA molecule (H/L), DNA samples were electrophoresed. The expected sizes of the HC, LC, and H/L products visualized on the gel were approximately 730, 690, and 1,390 base pairs, respectively.

Figures 24A and 24B illustrate the major features of the bacterial expression vector Lambda ZAP II Modified $V_H$ (Modified ImmunoZAP H) ($V_H$-expression vector) (IZ H). The amino acids encoded by the synthetic DNA sequence from Figure 24A is shown along with the $T_3$ polymerase promoter from Lambda ZAP II. The orientation of the insert in Lambda ZAP II is as presented. The insert was modified by the elimination of the Sac I site between the $T_3$ polymerase and Not I site and by the change of amino acids at the 5' end of the heavy chain from QVKL to QVQL (alysine residue was changed to a glutamine residue). The $V_H$ and $V_L$ DNA homologs were inserted into the Xho I and Xba I cloning sites of the phagemid as described in Figure 26 and shown in Figure 24B. The modifications were made to create a fusion-PCR library from hybridoma RNA, to overcome decreased efficiency of secretion of positively charged amino acids in the amino terminus of the protein. Inouye et al., Proc. Natl. Acad. Sci., USA, 85:7685-7689 (1988), and to make the $V_L$ Sac I cloning site a unique restriction site.

Figures 25A and 25B illustrate the sequences of the synthetic DNAs inserted into Lambda ZAP to produce Lambda Zap II $V_H$ (ImmunoZAP H) (25A) and Lambda Zap $V_L$ (ImmunoZAP L) (25B) expression vectors. The various features

required for these vectors to express the $V_H$ and $V_L$-coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., J. Biol. Chem., 255:27, 1980, and various restriction enzyme sites used to operatively link the $V_H$ and $V_L$ homologs to the expression vector. The $V_H$ expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions of the heavy chain ($V_H$ Backbone). This $V_H$ Backbone is just upstream and in the proper reading frame as the $V_H$ DNA homologs that are operatively linked into the Xho I and Spe I restriction sites. The $V_L$ DNA homologs are operatively linked into the $V_L$ sequence (25B) at the Sac I and Xba I restriction enzyme sites.

Figure 26 illustrates the major features of the bacterial expression vector Lambda Zap II $V_H$ (ImmunoZAP H) ($V_H$- expression vector). The amino acids encoded by the synthetic DNA sequence from Figure 25A is shown at the top along with the $T_3$ polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The $V_H$ DNA homologs were inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The $V_H$ DNA homologs were inserted into the Xho I and Spe I restriction enzyme sites. The read through transcription produces the decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 27 illustrates the major features of the bacterial expression vector Lambda Zap II $V_L$ (ImmunoZAP L) ($V_L$ expression vector). The amino acids encoded by the synthetic DNA sequence shown in Figure 25B is shown at the top along with the $T_3$ polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The $V_L$ DNA homologs are inserted into the Sac I and Xba I cloning sites of the phagemid as described in Figure 26.

Figure 28 illustrated an autoradiogram showing signals obtained from human phage clones. Approximately 100 lambda phage were spotted onto E. coli lawns, creating plaques that were overlaid with nitrocellulose filters previously soaked in 10 mM isopropylbeta-D-thiogalacto-pyranoside (IPTG) to induce Fab expression. Following overnight incubation, the filters were reacted with $^{125}I$-tetanus toxoid probe. After washing, the filters were exposed to X-ray film. The column on the right represents the parental clones that were selected form a combinatorial library. Mullinax et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990). The column on the left represents clones that were generated by amplifying, the combinatorial lambda clone DNA with the $V_H$ and $C_L$' outside primers, $C_H1$' and $V_L$ inside primers, followed by recloning in the modified ImmunoZAP H vector. Clone 7G1 is a negative control which expresses a Fab that does not react with tetanus toxoid. Clones 10C1 and 6C12 both produce Fabs that react with tetanus toxoid. IZ H is the modified heavy chain ImmunoZAP H vector without an insert.

## Detailed Description of the Invention

### A. Definitions

As used herein, the following terms have the following menings unless expressly stated to the contrary:

Nucleotide: a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3' or 5' position of the pentose it is referred to as a nucleotide.

Base Pair (bp): a pairing (by hydrogen bonding) of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule. In RNA, uracil (U) is substituted for thymine.

Nucleic Acid: a polymer of nucleotides, either single or double stranded.

Gene: a nucleic acid whose nucleotide sequence codes for an RNA or polypeptide. A gene can be either RNA or DNA.

Complementary Bases: nucleotides that normally pair up when DNA or RNA adopts a double stranded configuration.

Complementary Nucleotide Sequence: a sequence of nucleotides in a single-stranded molecule of DNA or RNA that is sufficiently complementary to that on another single strand to specifically hybridize to it with consequent hydrogen bonding.

Conserved: a nucleotide sequence is conserved with respect to a preselected (reference) sequence if it non-randomly hybridizes to an exact complement of the preselected sequence.

Hybridization: the pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex by the establishment of hydrogen bonds between complementary base pairs. It is a specific, i.e. non-random, interaction between two complementary polynucleotides that can be competitively inhibited.

Nucleotide Analog: a purine or pyrimidine nucleotide that differs structurally from A, T, G, C, or U, but is sufficiently similar to substitute for the normal nucleotide in a nucleic acid molecule.

DNA Homolog: is a nucleic acid having a preselected conserved nucleotide sequence and a sequence coding for a receptor capable of binding a preselected ligand.

Receptor: A receptor is a molecule, such as a protein, glycoprotein and the like, that can specifically (non-randomly) bind to another molecule.

Antibody: The term antibody in its various grammatical forms is used herein to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an

antibody combining site or paratope. Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v).

Antibody Combining Site: An antibody combining site is that structural portion of an antibody molecule comprised of a heavy and light chain variable and hypervariable regions that specifically binds (immunoreacts with) an antigen. The term immunoreact in its various forms means specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole antibody molecule or a portion thereof.

Monoclonal Antibody: The phrase monoclonal antibody in its various grammatical forms refers to a population of antibody molecules that contains only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different antigen, e.g., a bispecific monoclonal antibody.

Upstream: In the direction opposite to the direction of DNA transcription, and therefore going from 5' to 3' on the non-coding strand, or 3' to 5' on the mRNA.

Downstream: Further along a DNA sequence in the direction of sequence transcription or read out, that is traveling in a 3'- to 5'-direction along the non-coding strand of the DNA or 5'- to 3'-direction along the RNA transcript.

Cistron: Sequence of nucleotides in a DNA molecule coding for an amino acid residue sequence.

Stop Codon: Any of three codons that do not code for an amino acid, but instead cause termination of protein

synthesis. They are UAG, UAA and UGA. Also referred to as a nonsense or termination codon.

Leader Polypeptide: A short length of amino acid sequence at the amino end of a protein, which carries or directs the protein through the inner membrane and so ensures its eventual secretion into the periplasmic space and perhaps beyond. The leader sequence peptide is commonly removed before the protein becomes active.

Reading Frame: Particular sequence of contiguous nucleotide triplets (codons) employed in translation. The reading frame depends on the location of the translation initiation codon.

Inside Primer: An inside primer is a polynucleotide that has a priming region located at the 3' terminus of the primer which typically consists of 15 to 30 nucleotide bases. The 3' terminal-priming portion is capable of acting as a primer to catalyze nucleic acid synthesis. The 5'-terminal priming portion comprises a non-priming portion.

Outside Primer: An outside primer comprises a 3'-terminal priming portion and a portion that may define an endonuclease restriction site which is typically located in a 5'-terminal non-priming portion of the outside primer.

Fusion Polynucleotide Amplification: refers to *in vitro* techniques of generating a multiple complementary copies of a nucleic acid template which comprises nucleotide sequences which have been randomly combined to give a combined nucleic sequence. These techniques typically employ complementary primers which hybridize to the template and are extended in a primer extension reaction. The polyumerase chain reaction (PCR) techniques described herein comprise a preferred method of nucleotide sequence amplifications. Generation and amplification of a combined nucleotide sequence using fusion PCR is further described herein.

Vector: As used herein, the term "vector" refers to a nucleic acid molecule capable to transporting between different genetic environments another nucleic acid to which it has been operatively linked. One type of preferred vector is an episome, i.e., a nucleic acid molecule capable of extra-chromosomal replication. Other suitable vectors include plasmid and cosmid vectors and phage, especially bateriophage such as lambda. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors".

B. Methods

Until this invention, genetic engineers typically dealt with the expression of a single gene or family (or population) of genes, one at a time. The expression of a family of genes in a vector is generally referred to as a "gene library." Each member of the library will normally contain a different gene or DNA sequence. However, the vector portion of such a vector-gene fusion is typically identical from member to member. (Maniatis et al., supra). Individual members within the library may often be, and typically are, amplified before screening to identify and isolate a desired member. Amplification occurs so that each library member grows as a bacterial colony (for plasmid libraries) or phage plaques (for bacteriophage libraries, such as lambda). These amplified members are usually referred to as "clones," since each colony or plaque is made up of many identical host cells or phage particles.

The search for a particular clone containing a single gene or DNA sequence of interest can be accomplished in many different ways. The clone may be identified because its vector-gene specifically hybridizes with a nucleic acid probe. It may also be identified by expression of an

RNA species that can be identified, for example by nucleic acid hybridization. The RNA species may, furthermore, be translated into a protein, typically by the host cell, that may be identified, for example, by reactivity with an antibody probe. Alternatively, the protein may be recognized because it binds a substrate, or catalyzes a reaction, or allows the host cell to survive under selective conditions, and so on.

Described herein are libraries in which two or more families (or populations) of genes are expressed in a vector or a host cell in such a way that the gene combinations are randomly represented and subsequently detected on the basis of some property or characteristic in the event that a particular combination of one member from a first gene family and one gene from a one or more other gene families are combined in a vector host cell. For example, in the general case if there are "i" members of the gene family "A" and "j" members of the gene family "B", there will be (i) x (j) combinations of selected gene members A and B in the randomly created vector-gene population. If there are three gene families, A, B, and C, and a vector is made containing one member from each of the three gene families, the total number of combinations of genes will be the product of the number of A genes times the number of B genes times the number of C genes. Thus, methods are provided wherein at least two genes may randomly be combined, preferably on the same vector molecule, having been identified within a population of vectors containing other combinations of different genes from the same two or more gene families. This approach may be broadly accomplished by means other than recombination, for example, the use of a vector having at least two independent insertion sites for two foreign genes or inserting in a vector a nucleotide sequence comprising nucleotide sequences from each gene family. The recombination of at least two separate library populations to make a combinatorial population, for example, using a

common restriction site or site-directed recombination systems, is also contemplated.

Thus, in addition to the above-described methods, the invention also provides for vectors having characteristics and sequences useful for the preparation of combinatorial vectors encoding random DNA sequences from two or more gene families. Such vectors include plasmids and phage containing common restriction sites or sequences enabling the in vivo recombination of said DNA sequences from said gene families.

The flp site-specific recombination of S. cervisiae has been described in Cox, Chapter 13 in "Genetic Recombination," eds. R. Kucherlapati and G. Smith (American Society for Microbiology 1988). Within a sixty-five bp region identified as the recombination site and designated FRT (flp recombination target), there are several prominent structural features. The most important are a set of three bp repeats. The second and third repeats are separated by one bp and are in the same orientation. The first repeat is inverted with respect to the other two and is separated from the second repeat by an eight bp spacer. The first repeat also has a one bp mismatch relative to the first two. Deletion analysis has demonstrated that the third repeat is unnecessary for recombination in vitro, although it may have a slight effect on the reaction in vivo. Additional deletions indicate that most, but not all, of the first and second repeats (those flanking the spacer) are required. While deletion of three bp from the distal ends of one or both of these repeats has no detectable effect on the reaction, further deletion leads to a gradual reduction in site function, with complete loss of site function occurring (in vitro) with deletions of eight bp or more from either end. The minimal site required for a full function in vitro is therefore relatively small (approximately 28 bp including the spacer and the proximal 10 bp of each flanking repeat). Accordingly, it will be seen that the full,

intermediate, or minimal FRT sequences can be utilized to accomplish flp-mediated site-specific recombination.

The lambda phage attachment site is responsible for integration of lambda into the host chromosome. It also acts as a hot spot of recombination and lytic crosses between wild lambda chromosomes. As in lambda, in P1 phage a site-specific cross over site, loxP acts as a hot spot of recombination. This site is recognized by the P1 cre protein, a known site-specific protein. The site-specific recombination system is responsible for the rare integration of P1 into the host chromosome. The cre-lox system of bacteriaphage P1 is also useful for the site-specific recombination contemplated by the invention described and claimed herein.

A transposon can jump from one vector to another vector or from a vector to a bacterial chromosome. Different transposons having different inverted repeat sequences and carrying, for example, different drug-resistance genes, can be used to carry out the desired random combination of genes as described herein either in vivo or in vitro. The transposon may, but need not, also contain a sequence encoding the transposase enzyme which catalyzes the "hop." Various suitable transposon systems have been described in the literature. (See, Mobile DNA, Douglas E. Berg and Martha M. Howe, eds., American Society for Microbiology, Washington, D.C., 1989). One suitable transposon system is the gamma-delta transposon system which has been isolated from E. Coli.

Thus, in addition to restriction digestion and ligation, use of flp type recombination systems, and homologous recombination, a transposon system can also be used to integrate a light (or heavy) antibody chain clone into a heavy (or light) antibody chain clone. For example, this can be accomplished by flanking the light chain expression and cloning region with transposon terminal sequences. A library constructed in this light chain vector could be used to co-infect bacteria with

clones from the heavy chain library. The light chain inserts between the terminal sequences would hop from the light chain lambda phage vector into other DNA sequences in the presence of transposase activity. Selection for hopping into the heavy chain clone can be accomplished by placing a selectable marker within the light chain, positioned between the transposon hopping sequences. Subsequently, phage recovered from the co-infected culture is plated with a strain enabling selection for the heavy chain vector and for the light chain marker gene. Because this second plating is performed under conditions of a high cell to phage ratio, only one lambda phage will typically be introduced into each cell. The lambda phage should grow only if the phage contains genes from both the heavy and light chain clones; most efficiently resulting from the transposon hop. If the hop occurs in the essential genes of the heavy chain clone, the phage will not grow. Only phage containing the transposon in the proper position within the heavy chain will grow. A collection of these clones comprises a library of combinatorial heavy and light chain antibody clones.

According to one aspect of the present invention, fusion PCR is used to generate two PCR-amplified DNA fragments, each of which have one of their ends modified by directed mispriming so that those ends share regions of complementarity, i.e., cohesive termini. When the two fragments are mixed, denatured and reannealed in a PCR cycle, the cohesive termini on two strands hybridize to form an "overlapping" DNA duplex that is internally primed. The subsequent PCR cycle primer-extends the non-overlapping regions to form a hybride DNA molecule that is dicistronic. See Figure 21.

PCR amplification methods are described in detail in U.S. Patent Nos. 4,863,192, 4,683,202, 4,800,159, and 4,965,188, and at least in several texts including "PCR Technology: Principles and Applications for DNA Amplification", H. Erlich, ed., Stockton Press, New York

(1989); and "PCR Protocols" A Guide to Methods and Applications", Innis et al., eds., Academic Press, San Diego, California (1990).

Thus, in one aspect of the present invention, fusion PCR is used to produce a library of dicistronic DNA molecules ocntaining upstream and downstream cistrons wherein first and second PCR amplification products are produced using respective first and second PCR primer pairs. The first PCR primer pair comprises a first polypeptide outside primer and a first polypeptide inside primer. Similarly, the second PCR primer pair comprises a second polypeptide outside primer and a second polypeptide inside primer. The first and second polypeptide inside primers contain complementary 5'-terminal sequences that allow their DNA complements to hybridize and form an internally-primed duplex having 3'-overhanging termini. The internally-primed duplex is then subjected to primer extension reaction conditions to produce a double stranded, dicistronic DNA having substantially blunt or blunt ends. The dicistronic DNA is then PCR amplified using the outside primers as a PCR primer pair.

The dicistronic DNA molecule comprises two amino acid residue-coding sequences on the same strand separated by at least one stop codon and at least one signal sequence necessary for translation of the downstream cistron, such as a translation initiation codon, ribosome binding site, and the like. Thus, the upstream and downstream cistrons of the dicistronic DNA molecule are operatively linked by a cistronic bridge. The cistronic bridge comprises the genetic elements necessary to terminate translation of the upstream cistron and initiate translation of the downstream cistron. For instance, the coding strand of the bridge codes for one or more stop codons, preferably two, in the same translational reading frame as the upstream cistron. The cistronic bridge coding strand preferably also encodes a ribosome binding site for the dowstream cistron located downstream from the upstream cistron's

stop codon(s). Typically, the coding strand of the cistronic bridge will also encode a leader polypeptide segment in the same translational reading frame as the downstream cistron. When present, the nucleotide base sequence encoding the leader usually begins with an initiation codon located within an operative distance, i.e. is operatively linked, to the ribosome binding site.

The following discussion illustrates the use of fusion PCR to isolate a pair of $V_H$ and $V_L$ genes from the immunoglobulin gene repertoire. This discussion is not to be taken as limiting, but rather as illustrating an application of creating a novel phenotype by combining one member from each of two or more families of genes. The illustrated method can be used with other families of conserved genes which each for one unit of a dimeric receptor, whether obtained directly from a natural source, such naive or in vivo immunized cells, or from cells or one or more genes that have been treated or mutagenized in vitro. Generally, the method, combines the following elements:

1. Producing $V_H$ and $V_L$ gene repertoires.

2. Preparing sets of outside and inside polynucleotide primers for cloning polynucleotide segments containing immunoglobulin $V_H$ and $V_L$ region genes.

3. Preparing a library containing a plurality of different dicistronic DNA molecules, each containing a $V_H$ and a $V_L$ gene from the respective repertoires.

4. Expressing the dicistronic DNA molecules in suitable host cells.

5. Screening the polypeptides expressed by the dicistronic DNA molecules for the preselected activity, and segregating a dicistronic DNA molecules for the preselected activity, and segregating a dicistronic DNA molecule identified by the screening process.

The present invention also provides a novel method for screening variants of a parental clone or clones. If the parental clone or clones contain two

nucleotide sequences that, when expressed together, create a phenotype, then such nucleotide sequences can be altered to create populations of variants of such nucleotide sequences. If the two variant populations are coexpressed in a random fashion (that is with no correlation between the specific alterations made in the two different nucleotide sequences), then a combinatorial collection of such nucleotide sequence variants has been created. Such combinatorial collections may be screened for the presence of phenotypes that are unlike the parental clone or clones. Generally, the method combines the following elements:

1. Replicating a clone containing a nucleotide sequence under conditions that allow mutations to occur.

2. Replicating a second clone containing a second nucleotide sequence under conditions that allow mutations to occur.

3. Randomly combining and co-expressing the two mutated populations of nucleotide sequences.

4. Screening clones containing combinations of mutated nucleotide sequences for phenotypes that were not present in either parent clone.

Alternatively, the methods combine the following elements:

1. Replicating at least portions of two nucleotide sequences contained within a single clone under conditions that allow mutations to occur in either nucleotide sequence.

2. Allowing recombination events between the two nucleotide sequence populations to reassociate mutant nucleotide sequences to form new pairs of the two sequences that were not paired in the original mutated, replicated population.

3. Screening clones containing combinations of nucleotide sequences for phenotypes that were not present in the parent clone or in the mutant replicas of the parent clone.

For example, assume a parent clone containing two nucleotide sequences A and B is replicated under mutating conditions such that variant clones are formed:

Parent: A/B
Variant 1: A1/B
Variant 2: A/B1
Variant 3: A2/B1
Variant 4: A/B2
Variant 5: A3/B

However, within this mutated population, the combinations A1/B2, A2/B, A2/B2, A3/B1, and A3/B2, do not occur. If the mutant population (including some non-mutated parent clones) is allowed to recombine sequences A and B and their variants, then combinations such as A1/B2, A2/B etc. can be created. Such new combinations may express a desired phenotype that was not present in the parental or the variant population.

In one aspect, the present invention is related to methods for tapping the immunological repertoire by isolating from $V_H$-coding and $V_L$-coding gene repertoires genes coding for a heterodimeric antibody receptor capable of binding a preselected ligand. Generally, the method combines the following elements:

1. Isolating nucleic acids containing a substantial portion of the immunological repertoire.

2. Preparing polynucleotide primers for cloning polynucleotide segments containing immunoglobulin $V_H$ and $V_L$ region genes.

3. Preparing a gene library containing a plurality of different $V_H$ and $V_L$ genes from the repertoire.

4. Expressing the $V_H$ and $V_L$ polypeptides in a suitable host, including prokaryotic and eukaryotic hosts, on the same expression vector.

5. Screening the expressed polypeptides for the preselected activity, and segregating a $V_H$- and $V_L$-coding gene combination identified by the screening process.

In one aspect, the expressed phenotype produced by the methods by the present invention comprises a multimeric polypeptide product (i.e. a heterodimer, etc.) which assumes a conformation having a binding site specific for, as evidenced by its ability to be competitively inhibited, a preselected or predetermined ligand such as an antigen, enzymatic substrate and the like. In one embodiment, the multimeric polypeptide is an antibody that forms an antigen binding site which specifically binds to a preselected antigen to form an immunoreaction product (complex) having a sufficiently strong binding between the antigen and the binding site for the immunoreaction product to be isolated. The antibody typically has an affinity or avidity is generally greater than $10^5$-$M^{-1}$.

In another embodiment, a multimeric polypeptide produced according to the present invention is capable of binding a substrate and catalyzes the formation of a product from the substrate. While the topology of the ligand binding site of a catalying multimeric polypeptide is probably more important for its preselected activity than its affinity (association constant or pKa) for the substrate, the useful catalytic multimeric polypeptides typically have an association constant for the preselected substrate generally greater than $10^3$ $M^{-1}$, more usually greater than $10^5$ $M^{-1}$ or $10^6$ $M^{-1}$ and preferably greater than $10^7$ $M^{-1}$.

Preferably the multimeric polypeptide produced according to the present invention is heterodimeric and is therefore normally comprised of two different polypeptide chains, which together assume a conformation having a binding affinity, or association constant for the preselected ligand that is different, preferably higher, than the affinity or association constant of either of the polypeptides alone, i.e., as monomers. In a particularly preferred aspect, one or both of the different polypeptide chains is derived from the variable region of the light and heavy chains of an immunoglobulin. Typically, poly-

peptides comprising the light ($V_L$) and heavy ($V_H$) variable regions are employed together for binding the preselected ligand.

A $V_H$ or $V_L$ produced by the methods of the subject invention can be active in monomeric as well as multimeric forms, either homomeric or heteromeric, preferably heterodimeric. A $V_H$ and $V_L$ ligand binding polypeptide produced by the present invention can be advantageously combined in a heterodimer (antibody molecule) to modulate the activity of either or to produce an activity unique to the heterodimer. The individual ligand binding polypeptides will be referred to as $V_H$ and $V_L$ and the heterodimer will be referred to as an antibody molecule.

However, it should be understood that a $V_H$ binding polypeptide may contain in addition to the $V_H$, substantially all or a portion of the heavy chain constant region. A $V_L$ binding polypeptide may contain, in addition to the $V_L$, substantially all or a portion of the light chain constant region. A heterodimer comprised of a $V_H$ binding polypeptide containing a portion of the heavy chain constant region and a $V_L$ binding containing substantially all of the light chain constant region is termed a Fab fragment. The production of a Fab can be advantageous in some situations because the additional constant region sequences contained in a Fab as compared to a $F_v$ could stabilize the $V_H$ and $V_L$ interaction. Such stabilization could cause the Fab to have higher affinity for antigen. In addition the Fab is more commonly used in the art and thus there are more commercial antibodies available to specifically recognize a Fab.

The individual $V_H$ and $V_L$ polypeptides may be produced in lengths equal or substantially equal to their naturally occurring lengths. However, the individual $V_H$ and $V_L$ polypeptides will generally have fewer than 125 amino acid residues, more usually fewer than about 120 amino acid residues, while normally having greater than 60 amino acid residues, usually greater than about 95 amino acid

residues, more usually greater than about 100 amino acid residues. Preferably, the $V_H$ will be from about 110 to about 125 amino acid residues in length while $V_L$ will from about 95 to about 115 amino acid residues in length.

The amino acid residue sequences of the polypeptides will vary widely, depending upon the particular idiotype involved. Usually, there will be at least two cysteines separated by from about 60 to 75 amino acid residues and joined by a disulfide bond. The polypeptides produced by the subject invention will normally be substantial copies of idiotypes of the variable regions of the heavy and/or light chains of immunoglobulins, but in some situations a polypeptide may contain random mutations in amino acid residue sequences in order to advantageously improve the desired activity.

In some situations, it is desireable to provide for covalent cross linking of the $V_H$ and $V_L$ polypeptides, which can be accomplished by providing cysteine resides at the carboxyl termini. The polypeptide will normally be prepared free of the immunoglobulin constant regions, however a small portion of the J region may be included as a result of the advantageous selection of DNA synthesis primers. The D region will normally be included in the transcript of the $V_H$.

In other situations, it is desirable to provide a peptide linker to connect the $V_L$ and the $V_H$ to form a single-chain antigen-binding protein comprised of a $V_H$ and a $V_L$. This single-chain antigen-binding protein would be synthesized as a single protein chain. Such a single-chain antigen binding proteins have been described by Bird et al., Science, 242:423-426 (1988). The design of suitable peptide linker regions is described in U.S. Patent No. 4,704,692 by Robert Landner.

Such a peptide linker may be designed as part of the nucleic acid sequences contained in the expression vector. The nucleic acid sequences coding for the peptide linker would be between the $V_H$ and $V_L$ DNA homologs and the

restriction endonuclease sites used to operatively link the $V_H$ and $V_L$ DNA homologs to the expression vector.

Such a peptide linker also may be coded for nucleic acid sequences that are part of the polynucleotide primers used to prepare the various gene libraries. The nucleic acid sequence coding for the peptide linker can be made up of nucleic acids attached to one of the primers or the nucleic acid sequence coding for the peptide linker may be derived from nucleic acid sequences that are attached to several polynucleotide primers used to create the gene libraries.

Typically the C terminus region of the $V_H$ and $V_L$ polypeptides will have a greater variety of the sequences than the N terminus and, based on the present strategy, can be further modified to permit a variation of the normally occurring $V_H$ and $V_L$ chains. A synthetic polynucleotide and be employed by vary one or more amino in an hypervariable region.

1. Isolation Of A Gene Repertoire

According to one aspect of the present invention, a gene repertoire useful in the methods the present invention contains at least $10^3$, preferably at least $10^4$, more preferably at least $10^5$, and most preferably at least $10^7$ different consderved genes. Methods for evaulating the diversity of a repertoire of conserved genes are well known to one skilled in the art.

Various well known methods can be employed to produce a useful gene repertoire. For example, to prepare a composition of nucleic acids containing a substantial portion of the immunological gene repertoire, a source of genes coding for the $V_H$ and/or $V_L$ polypeptides is required. Preferably the source will be heterogeneous population of antibody producing cells, i.e., B lymphocytes (B cells), preferably rearranged B cells such as those found in the circulation or spleen of a vertebrate. (Rearranged B cells are those in which immunoglobulin gene transloca-

tion, i.e., rearrangement, has occurred as evidenced by the presence in the cell of mRNA with the immunoglobulin gene V, D and J region transcripts adjacently located thereon.)

In some cases, it is desirable to bias the repertoire for a preselected activity, such as by using as a source of nucleic acid cells (source cells) from vertebrates in any one of various stages of age, health and immune response. For example, repeated immunization of a healthy animal prior to collecting rearranged B cells results in obtaining a repertoire enriched for genetic material producing a ligand binding polypeptide of high affinity. See, e.g. Mullinax et al., Proc. Nat. Acad. Sci. (USA) 87:8095-8099 (1990). Conversely, collecting rearranged B cells from a healthy animal whose immune system had not been recently challenged results in producing a repertoire that is not biased towards the production of high affinity $V_H$ and/or $V_L$ polypeptides.

It should be noted the greater the genetic heterogeneity of the population of cells for which the nucleic acids are obtained, the greater the diversity of the immunological repertoire that will be made available for screening according to the method of the present invention. Thus, cells from different individuals of different strains, races or species can be advantageously combined to increase the heterogeneity (diversity) of the repertoire.

Thus, in one preferred embodiment, the source cells are obtained from a vertebrate, preferably a mammal, which has been immunized or partially immunized with an antigenic ligand (antigen) against which activity is sought, i.e., a preselected antigen. The immunization can be carried out conventionally. Antibody titer in the animal can be monitored to determine the stage of immunization desired, which stage corresponds to the amount of enrichment or biasing of the repertoire desired. Partially immunized animals typically receive only one immunization

and cells are collected therefrom shortly after a response is detected. Fully immunized animals display a peak titer, which is achieved with one or more repeated injections of the antigen into the host mammal, normally at 2 to 3 week intervals. Usually three to five days after the last challenge, the spleen is removed and the genetic repertoire of the spleenocytes, about 90% of which are rearranged B cells, is isolated using standard procedures. See, Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, NY.

Nucleic acids coding for $V_H$ and $V_L$ polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, most preferably from IgM and IgG, producing cells.

Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned as a diverse population are well known in the art. See for example Herrmann et al., Methods In Enzymol., 152:180-183, (1987); Frischauf, Methods In Enzymol., 152:180-190 (1987); Frischauf, Methods In Enzymol., 152:190-199 (1987); and DiLella et al., Methods In Enzymol., 152:199-212 (1987). (The teachings of the references cited herein are hereby incorporated by reference.)

The desired gene repertoire can be isolated from either genomic material containing the gene expressing the variable region or the messenger RNA (mRNA) which represents a transcript of the variable region. The difficulty in using the genomic DNA from other than non-rearranged B lymphocytes is in juxtaposing the sequences coding for the variable region, where the sequences are separated by intervening regions. The DNA fragment(s) containing the proper variable regions must be isolated, the intervening regions excised, and the variable regions then spliced in the proper order and in the proper orientation. For the most part, this will be difficult, so that the alternative technique employing rearranged B cells will be the method of choice because the V, D and J immunoglobulin gene

regions have translocated to become adjacent, so that the sequence is continuous for the variable regions.

Where mRNA is utilized the cells will be lysed under RNase inhibiting conditions. In one embodiment, the first step is to isolate the total cellular mRNA by hybridization to an oligo-dT cellulose column. The presence of mRNAs coding for the heavy and/or light chain polypeptides can then be assayed by hybridization with DNA single strands of the appropriate genes. Conveniently, the sequences coding for the constant portion of the $V_H$ and $V_L$ can be used as polynucleotide probes, which sequences can be obtained from available sources. See for example, Early and Hood, Genetic Engineering, Setlow and Hollaender, eds., Vol. 3, Plenum Publishing Corporation, New York, (1981), pages 157-188; and Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, MD, (1987). Exemplary methods for producing $V_H$ and $V_L$ gene repertoires are described in PCT Application No. PCT/US 90/02836 (International Publication No. WO 90/14430).

In preferred embodiments, the preparation containing the total cellular mRNA is first enriched for the presence of $V_H$ and/or $V_L$ coding mRNA. Enrichment is typically accomplished by subjecting the total mRNA preparation or partially purified mRNA product thereof to a primer extension reaction employing a polynucleotide synthesis primer of the present invention.

According to another aspect of the present invention, a gene repertoire may be generated from one or a few nucleotide sequences by replicating those sequences under mutagenesis conditions so that a plurality of different nucleotide sequences or genes may be generated. Suitable mutagenesis conditions are known to those skilled in the art.

2. Preparation Of Polynucleotide Primers

The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than 3. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

The term "primer" as used herein refers to a polynucleotide whether purified from a nucleic acid restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase, reverse transcriptase and the like, and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency, but may alternatively be stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is a polydeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agents for polymerization. The exact lengths of the primers will depend on many factors, including temperature and the source of primer. For example, depending on the complexity of the target sequence, a polynucleotide primer typically contains 15 to 25 or more nucleotides, although it can contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template.

The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be synthesized or amplified. This means that the primer must be sufficiently comple-

mentary to nonrandomly hybridize with its respective template strand. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such noncomplementary fragments typically code for an endonuclease restriction site. Alternatively, noncomplementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarily with the sequence of the strand to be synthesized to amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions.

Primers of the present invention may also contain a DNA-dependent RNA polymerase promoter sequence or its complement. See for example, Krieg et al., Nucleic Acids Research, 12:7057-70 (1984); Studier et al., J. Mol. Biol., 189:113-130 (1986); and Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

When a primer containing a DNA-dependent RNA polymerase promoter is used, the primer is hybridized to the polynucleotide strand to be amplified and the second polynucleotide strand of the DNA-dependent RNA polymerase promoter is completed using an inducing agent such as E. coli DNA polymerase I, or the Klenow fragment of E. coli DNA polymerase. The starting polynucleotide is amplified by alternating between the production of an RNA polynucleotide and DNA plynucleotide.

Primers may also contain a template sequence or replication initiation site for a RNA-directed polymerase. Typical RNA-directed RNA polymerase include the QB replicase described by Lizardi et al. Biotechnology, 6:1197-1202 (1988). RNA-directed polymerases produce large numebrs of RNA strands from a small number of template RNA strands that contain a template sequence or replication

initiation site. These polymerases typically give a one million-fold amplification of the template strand, as has been described by Kramer et al., J. Mol. Biol., 89:7819-736 (1974).

The polynucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester on phosphodiester methods see Narang et al., Meth. Enzymol., 68:90, (1979); U.S. Patent No. 4,356,270; and Brown et al., Meth. Enzymol., 68:109, (1979).

The choice of a primer's nucleotide sequence depends on factors such as the distance on the nucleic acid from the region coding for the desired receptor, its hybridization site on the nucleic acid relative to any second primer to be used, the number of genes in the repertoire it is to hybridize to, and the like.

(a) Primers for Producing $V_H$ and $V_L$ DNA Homologs

$V_H$ and $V_L$ gene repertoires can be separately prepared prior to their use in the methods of the present invention. Repertoire preparation is typically done by primer extension (or other in vitro amplificaiton method), preferably by primer extension in a PCR format.

For example, to produce $V_H$-coding DNA homologs by primer extension, the nucleotide sequence of a primer is selected to hybridize with a plurality of immunoglobulin heavy chain genes at a site substantially adjacent to the $V_H$-coding region so that a nucleotide sequence coding for a functional (capable of finding) polypeptide is obtained. To hybridize to a plurality of different $V_H$-coding nucleic acid strands, the primer must be a substantial complement of a nucleotide sequence conserved among the different strands. Such sites include nucleotide sequences in the constant region, any of the variable region framework regions, preferably the third framework region, leader region, promoter region, J region and the like.

If the $V_H$-coding and $V_L$-coding DNA homologs are to be produced by polymerase chain reaction (PCR) amplification,

two primers must be used for each coding strand of nucleic acid to be amplified. The first primer becomes part of the nonsense (minus or complimentary) strand and hybridizes to a nucleotide sequence conserved among $V_H$ (plus) strands within the repertoire. To produce $V_H$ coding DNA homologs, first primers are therefore chosen to hybridize to (i.e. be complementary to) conserved regions within the J region, CH1 region, hinge region, $C_H2$ region, or $C_H3$ region of immunoglobulin genes and the like. To produce a $V_L$ coding DNA homolog, first primers are chosen to hybridize with (i.e. be complementary to) a conserved region with the J region or constant region of immunoglobulin light chain genes and the like. Second primers become part of the coding (plus) strand and hybridize to a nucleotide sequence conserved among minus strands. To produce the $V_H$-coding DNA homologs, second primers are therefore chosen to hybridize with a conserved nucleotide sequence at the 5' end of the $V_H$-coding immunoglobulin gene such as in that area coding for the leader or first framework region. It should be noted that in the amplification of both $V_H$- and $V_L$-coding DNA homologs, the conserved 5' nucleotide sequence of the second primer can be complementary to a sequence exogenously added using terminal deoxynucleotidyl transferase as described by Loh et al., Science 243:217-220 (1989). One or both of the first and second primers can contain a nucleotide sequence defining an endonuclease recognition site. The site can be heterologous to the immunoglobulin gene being amplified and typically appears at or near the 5' end of the primer.

(b) Inside and Outside Primers

In one embodiment, the present invention utilizes a set of polynucleotides that form inside primers comprised of an upstream inside primer and a downstream inside primer. Each of the inside primers has a priming region located at the 3'-terminus of the primer. The priming region is typically the 3'-most (3'-terminal) 15 to 30

nucleotide bases. The 3'-terminal priming portion of each inside primer is capable of acting as a primer to catalyze nucleic acid synthesis, i.e., initiate a primer extension reaction off its 3' terminus. One or both of the inside primers is further characterized by the presence of a 5'-terminal (5'-most) non-priming portion, i.e., a region that does not participate in hybridization to repertoire template.

In fusion PCR, each inside primer works in combination with an outside primer to amplify a target nucleic acid sequence. The choice of PCR primer pairs for use in fusion PCR as described herein is governed by the same considerations as previously discussed for choosing PCR primer pairs useful in producing gen repertoires. That is, the primers have a nucleotide sequence that is complementary to a sequence conserved in the repertoire. Useful $V_L$ and $V_H$ inside priming sequences are shown in Tables 1 and 2, respectively, below.

Table 1

3' Priming Portions of Various Inside $V_L$ Primers

| Seq. Id. No. | |
|---|---|
| (1)[1] | 5' GTGATGACCCACTCTCC 3' |
| (2) | 5' GTGATGACCCAGTCTCCA 3' |
| (3) | 5' GTTGTGACTCAGGAATCT 3' |
| (4) | 5' GTGTTGACGCAGCCGCCC 3' |
| (5) | 5' GTGCTCACCCAGTCTCCA 3' |
| (6) | 5' CAGATGACCCAGTCTCCA 3' |
| (7) | 5' GTGATGACCCAGACTCCA 3' |
| (8) | 5' GTCATGACCCAGTCTCCA 3' |
| (9) | 5' TTGATGACCCAAACTCAA 3' |
| (10) | 5' GTGATAACCCAGGATGAA 3' |

[1] Nucleotides sequences 1-10 are unique 5' primers for the amplification of kappa light chain variable regions.

Table 2

3' Priming Portions of Various Inside $V_H$ Primers

| Seq. Id. No. | |
|---|---|
| (11)[1] | 5' ACAAGATTTGGGCTC 3' |
| (12)[2] | 5' TGGGGTTTTGAGCTC 3' |
| (13)[3] | 5' GAGACAGTGACCGGGTTCCTTGGCCCCA 3' |
| (14)[4] | 5' TGGAATGGGCACATGCAG 3' |
| (15)[5] | 5' TTATCATTTACCCGGAGA 3' |
| (16)[6] | 5' AACGGTAACAGTGGTGCCTTGGCCCCA 3' |
| (17)[7] | 5' ACAATCCCTGGGCACAAT 3' |
| (18)[8] | 5' CACCTTGGTGCTGCTGGC 3' |
| (19)[9] | 5' ACAACCACAATCCCTGGGCACAATTTT 3' |
| (20)[10] | 5' ACAATCCCTGGGCACAAT 3' |
| (21)[11] | 5' GAGTTCACTAGTTGGGCACGGTGGGCA 3' |

[1] Unique 3' primer for human IgG1, 2, 3, and 4 F.2d.

[2] Unique 3' primer for human $V_H$ amplification.

[3] 3' primer for amplifying human heavy chain variable regions.

[4] 3' primer for amplifying the Fd region of mouse IgM.

[5] 3' primer located in the CH3 region of human IgG1 to amplify the entire heavy chain.

[6] Unique 3' primer for amplification of mouse $F_V$.

[7] Unique 3' primer for amplification of mouse IgG1 Fd.

[8] Unique 3' primer for amplification of VH including part of the mouse gamma 1 first constant region.

[9] Unique 3' primer for amplification of VH including part of mouse gamma 1 first constant region and hinge region.

[10] 3' primer for amplifying mouse Fd including part of the mouse IgG first constant region and part of the hinge region.

[11] 3'primer for amplifying human IgG1 Fd including part of the human IgG first constant region and part of the hinge region including the two cysteines which create the disulfide bridge for producing Fab'2 (the primer corresponds to Kabat number 241QQ to 247).

A preferred set of inside primers used herein has primers with complementary 5'-terminal non-priming regions, the complementary strands of which are capable of hybridizing to each other to form a duplex with 3' overhangs. The duplex encodes all or part of a double stranded cistronic bridge. That is, if the 3' overhangs of the duplex are filled in with complementary bases so as to define a double stranded DNA extending from the 3'-terminus of one of the inside primers to the 3'-terminus of the other of the inside primers, that double stranded DNA segment forms a sequence of nucleotides that operatively links the upstream and downstream cistrons for polycistronic expression. Thus, while each of the inside primers in a set contains only a portion of the sequence information necessary to form the double stranded cistronic bridge, the two inside primers in combination encode both the plus and minus strands of all or part of the bridge.

For example, one inside upstream primer can have a sequence that forms a portion of the plus strand of the bridge, and the other inside primer encodes the sequence, through complementarity, of the downstream portion of the plus strand.

In a preferred embodiment, the plus strand of the cistronic bridge contains, in the translational reading frame and from an upstream position to a downstream position, sequences coding for (i) at least one stop codon, preferably two, in the same reading frame as the upstream cistron, (ii) a ribosome binding site, and (iii) a polypeptide leader, the translation initiation codon of which is in the same reading frame as the downstream cistron. The stop codon is present to terminate translation of the upstream cistron. The ribosome binding site is present to initiate translation of the downstream cistron from the polycistronic mRNA.

The predicted amino acid residue sequences of two pelB gene product variants from Erwinia Carotova are shown

in Table 3. Lei, et al., supra. Amino Acid residue sequences for other leaders from E. coli useful in this invention are also listed in Table 3. Oliver, In Neidhart, F. C. (ed.), Escherichia coli and Salmonella Typhimurium, American Society for Microbiology, Washington, D. C., 1:56-69 (1987). These regions for the heavy chain are contained in the modified ImmunoZAP H expression vector. Mullinax, et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990).

Table 3
Leader Sequences

| Seq. Id. No. | Type | Amino Acid Residue Sequence |
|---|---|---|
| (22) | pelB[1] | MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeu LeuLeuAlaAlaGlnProAlaGlnProAlaMetAla |
| (23) | pelB[2] | MetLysSerLeuIleThrProIleAlaAlaGlyLeuLeu LeuAlaPheSerGlnTyrSerLeuAla |
| (24) | MalE[3] | MetLysIleLysThrGlyAlaArgIleLeuAlaLeuSer AlaLeuThrThrMetMetPheSerAlaSerAlaLeuAla LysIle |
| (25) | OmpF[3] | MetMetLysArgAsnIleLeuAlaValIleValProAla LeuLeuValAlaGlyThrAlaAsnAlaAlaGlu |
| (26) | PhoA[3] | MetLysGlnSerThrIleAlaLeuAlaLeuLeuProLeu LeuPheThrProValThrLysAlaArgThr |
| (27) | Bla[3] | MetSerIleGlnHisPheArgValAlaLeuIleProPhe PheAlaAlaPheCysLeuProValPheAlaHisPro |
| (28) | LamB[3] | MetMetIleThrLeuArgLysLeuProLeuAlaValAla ValAlaAlaGlyValMetSerAlaGlnAlaMetAlaVal Asp |
| (29) | Lpp[3] | MetLysAlaThrLysLeuValLeuGlyAlaValIleLeu GlySerThrLeuLeuAlaGlyCysSer |

[1] pelB from Erwinia carotovora gene
[2] pelB from Erwinia carotovora EC 16 gene
[3] leader sequences from E. coli

To achieve high levels of gene expression in E. coli, it is necessary to use not only strong promoters to generate large quantities of mRNA, but also ribosome binding sites to ensure that the mRNA is efficiently translated. In E. coli, the ribosome binding site includes an initiation codon (AUG) and a sequence 3- nucleotides long located 3 11 nucleotides upstream from the initiation codon [Shine et al., Nature, 254:34 (1975)]. The sequence, AGGAGGU, which is called the Shine-Dalgarno (SD) sequence, is complementary to the 3' end of E. coli 16S mRNA. Binding of the ribosome to mRNA and the sequence at the 3' end of the mRNA can be affected by several factors:

(i) The degree of complementarity between the SD sequence and 3' end of the 16S tRNA.

(ii) The spacing and possibly the DNA sequence lying between the SD sequence and the AUG [Roberts et al., Proc. Natl. Acad. Sci. USA, 76:760 (1979A); Roberts et al., Proc. Natl. Acad. Sci. USA, 76:5596 (1979B); Guarente et al., Science, 209:1428 (1980); and Guarente et al., Cell, 20:543 (1980).] Optimization is achieved by measuring the level of expression of genes in plasmids in which this spacing is systematically altered. Comparison of different mRNAs shows that there are statistically preferred sequences from positions -20 to +13 (where the A of the AUG is position 0) [Gold et al., Annu. Rev. Microbiol., 35:365 (1981)]. Leader sequences have been shown to influence translation dramatically (Roberts et al. 1979 a, b supra).

(iii) The nucleotide sequence following the AUG, which affects ribosome binding [Taniguchi et al., J. Mol. Biol., 118:533 (1978)].

Useful ribosome binding sites are shown in Table 4 below.

Table 4

Ribosome Binding Sites*

| Seq. Id. No. | | |
|---|---|---|
| 1. | (30) | 5' AAUCUUGGAGGCUUUUUUAUGGUUCGUUCU |
| 2. | (31) | 5' UAACUAAGGAUGAAAUGCAUGUCUAAGACA |
| 3. | (32) | 5' UCCUAGGAGGUUUGACCUAUGCGAGCUUUU |
| 4. | (33) | 5' AUGUACUAAGGAGGUUGUAUGGAACAACGC |

* Sequences of initiation regions for protein synthesis in four phage mRNA molecules are underlined.

AUG = initiation codon (double underlined)

1. = Phage $\phi$X174 gene-A protein
2. = Phage Q$\beta$ replicase
3. = Phage R17 gene-A protein
4. = Phage lambda gene-cro protein

It is preferred that the complementary (overlapping) region of the inside primers and the priming portion of the inside primers have about the same denaturation temperature, Td. The Td of a sequence can be estimated by the following formula: Td = 4(C+G) + 2(A+T), where C, G, A and T represent the respective number of cytosine, guanine, adenine and thymine bases in the sequence. A Td for the above-identified hybridizing region of about 45-55°C, preferably about 50°C, is preferred. Typically, overlapping regions in the range of about 15 to 20 nucleotides works well in conjunction with the priming regions in the range of 15-30 nucleotides.

The set of outside primers forms the termini of the dicistronic DNA molecule. The set of outside primers comprises an upstream outside primer and a downstream outside primer. The outside primers each comprise a 3'-terminal priming portion, and preferably a portion that defines an endonuclease restriction site. When present, the restriction site-defining portion is typically located in a 5'-terminal non-priming portion of the outside primer. The restriction site defined by the upstream outside primer is typically chosen to be one recognized by

a restriction enzyme that does not recognize the restriction site defined by the downstream outside primer, the objective being to be able to produce a dicistronic DNA having cohesive termini that are non-compelementary to each other and thus allow directional insertion into a vector.

Useful outside primer sequences are shown in Tables 5 and 6 below.

Table 5

Outside $V_H$ Primers

| Seq. Id. No. | |
|---|---|
| (34)[1] | 5'AGGTCCAGCTGCTCGAGTCTGG3' |
| (35) | 5'AGGTCCAGCTGCTCGAGTCAGG3' |
| (36) | 5'AGGTCCAGCTTCTCGAGTCTGG3' |
| (37) | 5'AGGTCCAGCTTCTCGAGTCAGG3' |
| (38) | 5'AGGTCCAACTGCTCGAGTCTGG3' |
| (39) | 5'AGGTCCAACTGCTCGAGTCAGG3' |
| (40) | 5'AGGTCCAACTTCTCGAGTCTGG3' |
| (41) | 5'AGGTCCAACTTCTCGAGTCAGG3' |
| (42)[2] | 5'AGGTGCAGCTGCTCGAGTCTGG3' |
| (43) | 5'AGGTGCAGCTGCTCGAGTCGGG3' |
| (44) | 5'AGGTGCAACTGCTCGAGTCTGG3' |
| (45) | 5'AGGTGCAACTGCTCGAGTCGGG3' |

[1] Nucleotide sequences 21-28 are unique 5' primers for the amplification of mouse $V_H$ genes.

[2] Nucleotide sequences 29-32 are unique 5' primers for amplification of nucleic acids coding for human variable regions.

Table 6

Outside $V_L$ Primers

| Seq. Id. No. | |
|---|---|
| (46)[1] | 5' ACGTCTAGATTCCACCTTGGTCCC 3' |
| (47)[2] | 5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3' |

(48)[3] 5' GCATTCTAGACTATTAACATTCTGTAGGGGC 3'
(49)[4] 5' GCAGCATTCTAGAGTTTCAGCTCCAGCTTGCC 3'
(50)[5] 5' CCGCCGTCTAGAACACTCATTCCTGTTGAAGCT 3'
(51)[6] 5' CCGCCGTCTAGAACATTCTGCAGGAGACAGACT 3'
(52)[7] 5' GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA 3'
(53)[8] 5' GCCGCTCTAGAACACTCATTCCTGTTGAA 3'
(54)[9] 5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3'
(55)[10] 5' GCATTCTAGACTATTATGAACATTCTGTAGGGGC 3'

[1] 3' primer for amplifying human kappa chain variable regions.

[2] 3' primer in human kappa light chain constant region.

[3] 3' primer in human lambda light chain constant region.

[4] Unique 3' primer for amplification of kappa light chain variable regions.

[5] Unique 3' primer for mouse kappa light chain amplification including the constant region.

[6] Unique 3' primer for mouse lambda light chain amplification including the constant region.

[7] Unique 3' primer for amplification of kappa light chain.

[8] Unique 3' primer for amplification of mouse kappa light chain.

[9] Unique 3' primer for kappa $V_L$ amplification.

[10] Unique 3' primer for human, mouse and rabbit lambda $V_L$ amplification.

3. Preparing a Gene Library

The strategy used for cloning, i.e., substantially reproducing the $V_H$ and/or $V_L$ genes contained within the isolated repertoire will depend, as is well known in the art, on the type, complexity, and purity of the nucleic acids making up the repertoire. Other factors include whether or not the genes are contained in one or a plurality of repertoires or populations and whether or not they are to be amplified and/or mutagnized.

a. Preparing $V_H$ and $V_L$ libraries

In one strategy, the object is to clone the $V_H$- and/or $V_L$-coding genes from a repertoire comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. The repertoire is subjected to a first primer extension reaction by treating (contacting) the repertoire with a first polynucleotide synthesis primer having a preselected nucleotide sequence. The first primer is capable of initiating the first primer extension reaction by hybridizing to a nucleotide sequence, preferably at least about 10 nucleotides in length and more preferably at least about 20 nucleotides in length, conserved within the repertoire. The first primer is sometimes referred to herein as the "sense primer" because it hybridizes to the coding or sense strand of a nucleic acid. In addition, the second primer is sometimes referred to herein as the "anti-sense primer" because it hybridizes to a non-coding or anti-sense strand to a nucleic acid, i.e., a strand complementary to a coding strand.

The PCR reaction is performed by mixing the PCR pair, preferably a predetermined amount thereof, with the nucleic acids of the repertoire, preferably a predetermined amount thereof, in a PCR buffer to form a first PCR admixture. The admixture is maintained under polynucleotide synthesizing conditions for a time period, which is typically predetermined, sufficient for the formation of a PCR reaction product, thereby producing a gene library containing a plurality of different $V_H$- and/or $V_L$-coding DNA homologs.

A plurality of first primer and/or a plurality of second primers can be used in each amplification, e.g., one species of first primer can be paired with a number of second primers to form several different primer pairs. Alternatively, an individual pair of first and second

primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and second primers can be combined to increase the diversity of the gene library.

In another strategy, the object is to clone the $V_H$- and/or $V_L$-coding gene from a repertoire by providing a polynucleotide complement of the repertoire, such as the anti-sense strand of genomic dsDNA or the polynucleotide produced by subjecting mRNA to a reverse transcriptase reaction. Methods for producing such complements are well known in the art. The complement is subjected to a primer extension reaction similar to the above-described second primer extension reaction, i.e., a primer extension reaction using a polynucleotide synthesis primer capable to hybridizing to a nucleotide sequence conserved among a plurality of different $V_H$-coding gene complements.

The primer extension reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess (for genomic nucleic acid, usually about $10^6$:1 primer:template) of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

The deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP are also admixed to the primer extension (polynucleotide synthesis) reaction admixture in adequate amounts and the resulting solution is heated to about 90°C-100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to room temperature, which is preferable for primer hybridization. To the cooled mixture is added an appropriate agent for inducing or catalyzing the primer extension reaction, and the reaction is allowed to occur under conditions known in the art. The synthesis reaction may occur at from room temperature up to a temperature above which the inducing agent no longer functions effi-

ciently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature is generally no greater than about 40°C.

The inducing agent may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli, DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be inducing agents, however, which initiate synthesis at the 5' end and proceed in the above direction, using the same process as described above.

The inducing agent also may be a compound or system which will function to accomplish the synthesis of RNA primer extension products, including enzymes. In preferred embodiments, the inducing agent may be a DNA-dependent RNA polymerase such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase. These polymerases produce a complementary RNA polynucleotide. The high turn overrate of the RNA polymerase amplifies the starting polynucleotide as has been described by Chamberlin et al., The Enzymes, ed. P. Boyer, PP. 87-108< Academic Press, New York (1982). Another advantage of T7 RNA polymerase is that mutations can be introduced into the polynucleotide synthesis by replacing a portion of cDNA with one or more mutagenic oligodeoxynucleotides (polynucleotides) and transcribing the partially-mismatched template directly as has been previously described by Joyce et al., Nucleic Acid Research, 17:711-722 (1989). Amplification systems

based on transcription have been described by Gingeras et al., in PCR Protocols, A Guide to Methods and Applications, PP. 245-252, Academic Press, Inc., San Diego, CA (1990).

If the inducing agent is a DNA-dependent RNA polymerase and therefore incorporates ribonucleotide triphosphates, sufficient amounts of ATP, CTP, GTP and UTP are admixed to the primer extension reaction admixture and the resulting solution is treated as described above.

The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which can be used in the succeeding steps of the process.

The first and/or second primer extension reaction discussed above can advantageously be used to incorporate into the multimeric polypeptide a preselected epitope useful in immunologically detecting and/or isolating a multimeric polypeptide. This is accomplished by utilizing a first and/or second polynucleotide synthesis primer or expression vector to incorporate a predetermined amino acid residue sequence into the amino acid residue sequence of the receptor.

After producing $V_H$- and/or $V_L$-coding DNA homologs for a plurality of different $V_H$- and/or $V_L$-coding genes within the repertoire, the homologs are typically amplified. While the $V_H$ and/or $V_L$-coding DNA homologs can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the DNA homologs by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a vector. In fact, in preferred strategies, the first and/or second primer extension reactions used to produce the gene library are the first and second primer extension reactions in a polymerase chain reaction.

PCR is typically carried out by cycling i.e., simultaneously performing in one admixture, the above described first and second primer extension reactions, each cycle comprising polynucleotide synthesis followed by

denaturation of the double stranded polynucleotides formed. Methods and systems for amplifying a DNA homolog are described in U.S. Patents No. 4,683,195 and No. 4,683,202, both to Mullis et al. Preferably, PCR is carried out by thermocycling i.e., repeatedly increasing and decreasing the temperature of a PCR reaction admixture within a temperature range whose lower limit is about 10°C to about 50°C and whose upper limit is about 90°C to about 100°C. The increasing and decreasing can be continuous, but is preferably phasic with time periods of relative temperature stability at each of temperatures favoring polynucelotide synthesis, denaturation and hybridization.

In preferred embodiments only one pair of first and second primers is used per amplification reaction. The amplification reaction products obtained from a plurality of different amplifications, each using a plurality of different primer pairs, are then combined.

However, the present invention also contemplated DNA homolog production via co-amplification (using two pairs of primers), and multiplex amplification (using up to about 8, 9 or 10 primer pairs).

The $V_H$- and $V_L$-coding DNA homologs produced by PCR amplification are typically in double-stranded form and have contiguous or adjacent to each of their termini a nucleotide sequence defining an endonuclease restriction site. Digestion of the $V_H$- and $V_L$-coding DNA homologs having restriction sites at or near their termini with one or more appropriate endonucleases results in the production of homologs having cohesive termini of predetermined specificity.

In preferred embodiments, the PCR process is used not only to amplify the $V_H$- and/or $V_L$-coding DNA homologs of the library, but also to induce mutations within the library and thereby provide a library having a greater heterogeneity. First, it should be noted that the PCR processes itself is inherently mutagenic due to a variety of factors well known in the art. Second, in addition to

the mutation inducting variations described in the above referenced U.S. Patent No. 4,683,195, other mutation inducing PCR variations can be employed. For example, the PCR reaction admixture, i.e., the combined first and second primer extension reaction admixtures, can be formed with different amounts of one or more of the nucleotides to be incorporated into the extension product. Under such conditions, the PCR reaction proceeds to produce nucleotide substitutions within the extension product as a result of the scarcity of a particular base. Similarly, approximately equal molar amounts of the nucleotides can be incorporated into the initial PCR reaction admixture in an amount to efficiently perform X number of cycles, and then cycling the admixture through a number of cycles in excess of X, such as, for instance, 2X. Alternatively, mutations can be induced during the PCR reaction by incorporating into the reaction admixture nucleotide derivatives such as inosine, not normally found in the nucleic acids of the repertoire being amplified. During subsequent in vivo amplification, the nucleotide derivative will be replaced with a substitute nucleotide thereby inducting a point mutation.

b. Preparing a Dicistronic DNA molecule Library

In one embodiment, a library of dicistronic DNA molecules containing upstream and downstream cistrons operatively linked by a cistronic bridge can be produced by the following steps:

(a) Subjecting a repertoire of first polypeptide genes (e.g., $V_H$-coding genes), to PCR amplification using first outside and first inside primers, i.e., a first PCR primer pair, to form a first primary PCR product.

(b) Subjecting a repertoire of second polypeptide genes (e.g., $V_L$-coding genes) to PCR amplification using second outside and second inside pirmers, i.e., a second PCR primer pair, to form a second primary PCR product.

(c) Hybridizing the first and second primary PCR products to form internally (self) primed duplexes, i.e., duplexes having 3'-hybridized and 5'-overhanging termini.

(d) Subjecting the internally-primed duplexes to primer extension reaction conditions to form double stranded duplexes having substantially blunt, preferably blunt, termini and a dicistronic strand containing the upstream and downstream cistrons linked by a cistronic bridge encoded by the inside primers. By "substantially blunt" is meant having no more than about one or two overhanging nucleotides. (Substantialloy blunt double stranded DNA is sometimes produced by primer overextension by Taq polymerase, usually by the addition of one or two terminal adenine residues.)

The $V_H$- and $V_L$-coding gene repertoires are comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. A repertoire is subjected to a PCR reaction as described in Section 3a hereinabove.

In preferred embodiments the ratio of gene molecules and their respective primers is as follows: about $1 \times 10^3$ $V_H$ gene molecules to about $1 \times 10^8$ outside $V_H$ gene molecules to about $1 \times 10^8$ outside $V_H$ primer molecules, about $1 \times 10^3$ $V_H$ gene molecules, to about $1 \times 10^7$ inside $V_H$ gene primer molecules, about $1 \times 10^3$ $V_L$ gene molecules to about $1 \times 10^8$ outside $V_L$ gene primer molecules, about $1 \times 10^4$ $V_L$ gene molecules to about $1 \times 10^7$ $V_L$ gene primer molecules. In more preferred embodiments, $10^4$ outside $V_H$ gene primer molecules and $10^3$ inside $V_H$ gene primer molecules are used for every $V_H$ gene molecule present in the PCR admixture. Similarly, $10^4$ outside $V_L$ gene primer molecules and $10^3$ $V_L$ gene molecule present in the PCR admixture. Thus, there is typically a 10 fold molar excess of outside primer to inside primer.

In the fusion PCR reaction, the gene repertoires are admixed with outside and inside primers, the outside primers being present in excess relative to the inside primers. The initial PCR thermocycles produce intermediate products having complementary termini from each of the first and second gene repertoires. That is, the end of one strand from one primary PCR product is capable of hybridizing with the complementary end from the other primary PCR product. The strands having the overlap at their 3' ends can act as primers for one another, i.e., from an internally primed duplex, and be extended by the polymerase to form the full length final product. The final product is then amplified by the set of outside primers, which act as a third PCR pair when the inside primers have been exhausted, to form a secondary PCR product. Typically the molar ratio of outside primers to inside primers is such that the inside primers are effectively exhausted within about 2 to about 12, preferably about 5, 6 or 7 thermocycles.

The PCR buffer also contains the deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and a polymerase, typically thermostable, all in adequate amounts for primer extension (polynucleotide synthesis) reaction. The resulting solution (PCR admixture) is heated to about 90°C - 100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to 54°C, which is preferably for primer hybridization. The synthesis reaction may occur at from room temperature up to a temperature above which the polymerase (inducing agent) no longer functions efficiently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature is generally no greater than about 40°C. An exemplary PCR buffer comprises the following: 50 mM KCl; 10 mM Tris-HCl; pH 8.3; 1.5 mM $MgCl_2$; 0.001% (wt/vol) gelatin, 200 $\mu$M dATP; 200 $\mu$M dTTP; 200 $\mu$M dCTP; 200 $\mu$M dGTP; and 2.5 units Thermus acquaticus DNA poly-

merase I (U.S. Patent No. 4,889,818) per 100 microliters of buffer.

After producing operatively linked $V_H$- and $V_L$-coding DNA homologs for a plurality of different $V_H$- and $V_L$-coding genes within the repertoires, the dicistronic DNA molecules are typically further amplified. While the dicistronic DNA molecules can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the molecules by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a vector. In fact, in preferred strategies, the first and second PCR reactions are performed in the same admixture that is subject to a multiplicity of PCR thermocycles where the outside primers are in molar excess. Preferably the number of PCR thermocycles is at least n+5, wherein n is the number of PCR thermocycles necessary to decrease by a factor of 10, and preferably exhaust, the number of inside primers by consumption in the formation of inside primer-primed products.

A diverse library of dicistronic DNA molecules having upstream and downstream cistrons can also be produced by combining, in a PCR buffer, double stranded $V_H$ and $V_L$ repertoires, $V_H$ and $V_L$ outside primers, and an inside primer having a 3'-teminal priming portion, a cistronic bridge coding portion, and a 5'-terminal inside primer-template (primer-coding) portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion has a nucleotide base sequence homologous (identical) to a protion of the primer extension product of the other of the outside primers. That is, the linking primer has terminal sequences homologous to sequences in both repertoires. The cistronic bridge coding portion codes for, either directly or through complementarily, at least one stop codon in the same reading frame as the upstream

cistron and sequences for the expression of the downstream cistron.

The dicistronic DNA molecules containing operatively linked $V_H$- and $V_L$-coding DNA homologs produced by PCR amplification are typically in double-stranded in form and may have contiguous or adjacent to each of their termini a nucleotide sequence defining an endonuclease restriction site. Digestion of the dicistronic DNA molecules having restriction sites at or near their temini with one or more appropriate endonucleases results in the production of DNA molecules having cohesive termini of predetermined specificity.

When individual PCR admixtures contain diverse gene repertoires the present invention produces many non-naturally occurring antibodies, i.e., combinations of $V_H$ and $V_L$ in a heterodimer. To take advantage of the mammalian immune system's capacity to select $V_H$ and $V_L$ combinations, the present invention also contemplates using fusion PCR to operatively link, and thereby recover, naturally occurring $V_H$ and $V_L$ combinations.

In certain preferred embodiments, a fusion PCR method is performed on repertoires comprising a plurality of substantially isolated cells containing genes coding for a heterodimeric receptor. For example, a plurality of PCR admixtures is formed, each of which contains (i) a sample of substantially isolated B lymphocytes from a mammal producing antibody molecules against a preselected antigen, (ii) a PCR buffer, and (iii) either the previously described $V_H$ and $V_L$ PCR primer pairs or the set of outside $V_H$ and $V_L$ PCR primers in combination with the linking primer(s), also as previously described. The plurality of PCR admixtures is then subjected to a multiplicity of PCR thermocycles as described herein.

By "substantially isolated" is meant a sample containing less than about 100 target cells, such as B lymphocytes, T cells, and the like. In preferred embodiments, the plurality of PCR admixtures contain only about

one cell. The cells are typically obtained from an individual mammal whose serum contains antibody molecules against the preselcted antigen. The collected cells are typically seeded, usually at densities in the range of 0.5 to 100 cells per unit volume, into a plurality of individual PCR vessels, such as microtiter plate wells and the like. Usually, the plurality of PCR admixtures is in the range of 800 to 1200, and preferably is about 1000, separate admixtures.

Typically, fewer cells are needed in each PCR admixture where the cells are obtained from individuals expressing a high serum antibody titer against the preselected antigen. For example, where B lymphocytes are obtained from an individual having a frequency of circulating B cells producing the antibody molecules of preselected specificity of 1/3000, each of about 800 to 1200 individual PCR admixtures need only contain about one B lymphocyte to result in isolation of the desired antibody. Where the circulating B cell frequency is in the range of 1/500,000, a density of about 100 cells per PCR admixture in each of about 800 to 1200 individual PCR admixtures will be needed before the process will result in isolation of the desired antibody.

In preferred embodiments, the PCR process is used not only to produce a library of dicistronic DNA molecules, but also to induce mutations within the library or to create diversity from a single parental clone and thereby provide a library having a greater heterogeneity as noted in Section 3a hereinabove.

4. Expression

A. Expressing the $V_H$ and/or $V_L$ DNA Homologs.

The $V_H$- and/or $V_L$-coding DNA homologs contained within the library produced by the above-described method can by operatively linked to a vector for amplification and/or expression.

The choice of vector to which a $V_H$- and/or $V_L$-coding DNA homolog is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., replication or protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. In preferred embodiments, the vector utilized includes a procaryotic replicon i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra chromosomally in a procaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, those embodiments that include a procaryotic replicon also include a gene whose expression confers a selective advantage, such as drug resistance, to a bacterial host transformed therewith. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

Those vectors that include a procaryotic replicon can also include a procaryotic promoter capable of directing the expression (transcription and translation) of the $V_H$- and/or $V_L$-coding homologs in a bacterial host cell, such as E. coli transformed therewith. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenience restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories, (Richmond, CA) and pPL and pKK223 available from Pharmacia, (Piscataway, NJ).

Promoters contain two highly conserved regions, one located about 10 bp (-10 region on Priberrow box) and the other about 35 bp (-35 region) upstream from the point at which transcription starts. These two regions typically determine promoter strength. In addition, the number of

nucleotides atht separate the conserved sequences is important for efficient promoter function. For example, 16 to 19 nucleotides typically separate the -10 and -35 regions, and changes in that psacing can change the efficiency of a promoter.

Promoters useful in this invention include Ptac $\phi$ 1.1A, $\phi$ 1.1B and $\phi$ 10, which are recognized by T7 polymerase. See U.S. Patent No. 4,946,786. Useful regulatable promoters include the E. coli lac promoter described in U.S. Patent No. 4,936,786 and the promoters for the temperature sensitive genes in U.S. Patent No. 4,806,471. See also U.S. Patent No. 4,711,845.

Expression vectors compatible with eukaryotic cells, preferably those compatible with vertebrate cells, can also be used. Eukaryotic cell expression vectors are well known in the art and are available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA homologue. Typical of such vectors are $pSV_L$ and pKSV-10 (Pharmacia), pBPV-1/PML2d (International Biotechnologies, Inc.), and pTDT1 (ATCC, No. 31255).

In preferred embodiments, the eukaryotic cell expression vectors used include a selection marker that is effective in an eukaryotic cell, preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, i.e., the neomycin phosphotransferase (neo) gene. Southern et al., J. Mol. Appl. Genet., 1:327-341 (1982).

The use of retroviral expression vectors to express the genes of the $V_H$ and/or $V_L$-coding DNA homologs is also contemplated. As used herein, the term "retroviral expression vector" refers to a DNA molecule that includes a promoter sequences derived from the long terminal repeat (LTR) region of a retrovirus genome.

In preferred embodiments, the expression vector is typically a retroviral expression vector that is preferably replication-incompetent in eukaryotic cells. The

construction and use of retroviral vectors has been described by Sorge et al., Mol. Cel. Biol., 41730-1737 1984).

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary cohesive termini can be engineered into the $V_H$- and/or $V_L$-coding DNA homologs during the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as previously discussed. The vector, and DNA homolog if necessary, is cleaved with a restriction endonuclease to produce termini complementary to those of the DNA homolog. The complementary cohesive termini of the vector and the DNA homolog are then operatively linked (ligated) to produce a unitary double stranded DNA molecule.

In preferred embodiments, the $V_H$-coding and $V_L$-coding DNA homologs of diverse libraries are randomly combined in vitro for polycistronic expression from individual vectors. That is, a diverse population of double stranded DNA expression vectors is produced wherein each vector expresses, under the control of a single promoter, one $V_H$-coding DNA homolog and one $V_L$-coding DNA homolog, the diversity of the population being the result of different $V_H$- and $V_L$-coding DNA homolog combinations.

Random combination in vitro can be accomplished using two expression vectors distinguished from one another by the location on each of a restriction site common to both. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein. In the first vector, the site is located between a promoter and a polylinker, i.e., 5' terminal (upstream relative to the direction of expression) to the polylinker by 3' terminal (downstream relative to the direction of expression). In the second vector, the polylinker is located between a promoter and the restriction site, i.e., the restriction site is located 3' terminal to the polylinker, and polylinker is located 3' terminal to the promoter.

In preferred embodiments, each of the vectors defines a nucleotide sequence coding for a ribosome binding and a leader, the sequence being located between the promoter and the polylinker, but downstream (3' terminal) from the shared restriction site if that site is between the promoter and polylinker. Also preferred are vectors containing a stop codon downstream from the polylinker, but upstream from any shared restriction site if that site is downstream from the polylinker. The first and/or second vector can also define a nucleotide sequence coding for a peptide tag. The tag sequence is typically located downstream from the polylinker but upstream from any stop codon that may be present.

In preferred embodiments, the vectors contain selectable markers such that the presence of a portion of that vector, i.e. a particular lambda arm, can be selected for or selected against. Typical selectable markers are well known to those skilled in the art. Examples of such markers are antibiotic resistance genes, genetically selectable markers, mutation suppressors such as amber suppressors and the like. The selectable markers are typically located upstream of the promoter and/or downstream of the second restriction site. In preferred embodiments, one selectable marker is located upstream of the promoter on the first vector containing the $V_H$-coding DNA homologs. A second selectable marker is located downstream of the second restriction site on the vector containing the $V_L$-coding DNA homologs. This second selectable marker may be the same or different from the first as long as when the $V_H$-coding vectors and the $V_L$-coding vectors are randomly combined via the first restriction site the resulting vectors containing both $V_H$ and $V_L$ and both selectable markers can be selected.

Typically the polylinker is a nucleotide sequence that defines one or more, preferably at least two, restriction sites, each unique to the vector, i.e., if it is on the first vector, it is not on the second vector.

The polylinker restriction sites are oriented to permit ligation of $V_H$- or $V_L$-coding DNA homologs into the vector in same reading frame as any leader, tag or stop codon sequence present.

Random combination is accomplished by ligating $V_H$-coding DNA homologs into the first vector, typically at a restriction site or sites within the polylinker. Similarly, $V_L$-coding DNA homologs are ligated into the second vector, thereby creating two diverse populations of expression vectors. It does not matter which type of DNA homolog, i.e., $V_H$ or $V_L$, is ligated to which vector, but it is preferred, for example, that all $V_H$-coding DNA homologs are ligated to either the first of second vector, and all of the $V_L$-coding DNA homologs are ligated to the other of the first or second vector. The members of both populations are then cleaved with an endonuclease at the shared restriction site, typically by digesting both populations with same enzyme. The resulting product is two diverse populations of restriction fragments where the members of one have cohesive termini complementary to the cohesive termini of the members of the other. The restriction fragments of the two populations are randomly ligated to one another, i.e., a random, interpopulation ligation is performed, to produce a diverse population of vectors each having a $V_H$-coding and $V_L$-coding DNA homolog located in the same reading frame and under the control of second vector's promoter. Of course, subsequent recombinations can be effected through cleavage at the shared restriction site, which is typically reformed upon ligation of members from the two populations, followed by subsequent religations.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the $V_H$- and/or $V_L$-coding DNA homologs, either separately or in combination. When coexpressed within the same organism, either on the same or the difference vectors, a functionally active Fv is produced. When the

$V_H$ and $V_L$ polypeptides are expressed in different organisms, the respective polypeptides are isolated and then combined in an appropriate medium to form a Fv. Cellular hosts into which a $V_H$- and/or $V_L$-coding DNA homolog-containing construct has been introduced are referred to herein as having been "transformed" or as "transformants".

The host cell can be either procaryotic or eucaryotic. Bacterial cells are preferred procaryotic host cells and typically are a strain of E. coli such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eucaryotic host cells include yeast and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is accomplished by methods that typically depend on the type of vector used. With regard to transformation of procaryotic host cells, see, for example, Cohen et al., Proceedings National Academy of Science, USA Vol. 69, P. 2110 (1972); and Maniatis et al., Molecular Cloning, a Laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). With regard to the transformation of vertebrate cells with retroviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); and Wigler et al., Proceedings National Academy of Sciences, USA, Vol. 76, P. 1373-1376 (1979).

b. Expressing the Dicistronic DNA Molecules

The dicistronic DNA molecules produced by the above-described method can be operatively linked to a vector for amplification and/or expression.

A variety of methods have been developed to oepratively link DNA to vectors via complemenetary cohesive termini. For instance, complementary cohesive termini can be engineered into the dicistronic DNA molecules during

the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as previously discussed. The dicistronic DNA molecule, and vector if necessary, is cleaved with a restriction endonuclease to produce termini complementary to htose of the vector. The complementary cohesive termini of the vector and the dicistronic DNA molecule are then operatively linked (ligated) to produce a unitary double stranded DNA molecule.

The present method produces a diverse population of double stranded DNA expression vectors wherein each vector expresses, under the control of a single promoter, one $V_H$-coding DNA homolog and one $V_L$-coding DNA homolog, the diversity of the populuation being the result of different $V_H$- and $V_L$-coding DNA homolog combination that occurs during the PCR reaction where both outside and both inside primers are present in effective amounts. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein.

In preferred embodiments, the vector defines a nucleotide sequence coding for a ribosome binding site and a leader, the sequence being located downstream from a promoter and upstream from a sequence ocding for apoly-peptide leader. In preferred embodiments, the vector contains a selectable marker such that the presence of a dicistronic DNA molecule of this invention inserted into the vector, can be selected. Typical selectable markers are well known to those skilled in the art. Examples of such markers are antibiotic resistance genes, genetically selectable markers, mutation suppressors such as amber supppressors and the like. The selectable markers are typically located upstream of the promoter.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the $V_H$- and $V_L$-coding DNA homologs. When coexpressed within the same organism, a functionally active heterodimeric receptor, such as an $F_v$, is produced.

Cellular hosts into which a $V_H$- and $V_L$-coding DNA homolog-containing constructu has been introduced are referred to herein as having been "transformed" or as "transformants".

The host cell can be either prokaryotic or eukaryotic. Bacterial cells are preferred prokaryotic host cells for library screening, and typically are a strain of E. coli such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eukaryotic host cells include yeast and mammalian cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is accomplished by methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen et al., Proc. Natl. Acad. Sci., USA, 69:2110 (1972); and Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY (1982). With regard to the transformation of vertebrate cells with retorviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); and Wigler et al., Proc. Natl. Acad. Sci., USA, 76:1373-1376 (1979).

5. Screening For Expression of $V_H$ and/or $V_L$ Polypeptides

Successfully transformed cells, i.e., cells containing a $V_H$- and/or $V_L$-coding DNA homolog or a dicistromic DNA molecule operatively linked to a vector, can be identified by any suitable well known technique for detecting the binding of a receptor to a ligand or the presence of a polynucleotide coding for the receptor, preferably its active site. Preferred screening assays are those where the binding of ligand by the receptor produces a detectable signal, either directly or indirectly. Such signals include, for example, the production of a complex,

formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, cells from a population subjected to transformation with a subject rDNA can be cloned to produce monoclonal colonies. Cells form those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a method such as that described Southern, J. Mol. Biol., 98:503 (1975) or Berent et al., Biotech. 3:208 (1985).

In addition to directly assaying for the presence of a $V_H$- and/or $V_L$-coding DNA homolog or a dicistronic DNA molecule, successful transformation can be confirmed by well known immunological methods, especially when the $V_H$ and/or $V_L$ polypeptides produced contain a preselected epitope. For example, samples of cells suspected of being transformed are assayed for the presence of the preselected epitope using an antibody against the epitope.

6. $V_H$- And/Or $V_L$-Coding Gene Libraries

According to one aspect, the present invention contemplates a gene library, preferably produced by a primer extension reaction or combination of primer extension reactions as described herein, containing at least about $10^3$, preferably at least about $10^4$ and more preferably at least about $10^5$ different $V_H$- and/or $V_L$-coding DNA homologs. The homologs are preferably in an isolated form, that is, substantially free of materials such as, for example, primer extension reaction agents and/or substrates, genomic DNA segments, and the like.

In preferred embodiments, a substantial portion of the homologs present in the library are operatively linked to a vector, preferably operatively linked for expression to an expression vector.

Preferably, the homologs are present in a medium suitable for in vitro manipulation, such as water, water containing buffering salts, and the like. The medium should be compatible with maintaining the activity of the homologs. In addition, the homologs should be present at

a concentration sufficient to allow transformation of a host cell compatible therewith at reasonable frequencies.

It is further preferred that the homologs be present in compatible host cells transformed therewith.

C. Expression Vectors

The present invention also contemplates various expression vectors useful in performing, inter alia, the methods of the present invention. Each of the expression vectors is a novel derivative of Lambda Zap vector.

1. Lambda Zap II

Lambda Zap II is prepared by replacing the Lambda S gene of the vector Lambda Zap with the Lambda S gene from the Lambda gt10 vector, as described in Example 6.

2. Lambda Zap II $V_H$

Lambda Zap II $V_H$ is prepared by inserting the synthetic DNA sequences illustrated in Figure 6A into the above-described Lambda Zap II vector. The inserted nucleotide sequence advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper initiation of mRNA translation into protein, and a leader sequence to efficiently direct the translated protein to the periplasm. The preparation of Lambda Zap II $V_H$ is described in more detail in Example 9, and its features illustrated in Figures 6A and 7.

3. Lambda Zap II $V_L$

Lambda Zap II $V_L$ is prepared as described in Example 12 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda Zap II $V_L$ are illustrated in Figure 8.

4. Lambda Zap II $V_L$ II

Lambda Zap II $V_L$ II is prepared as described in Example 11 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 10.

5. HCFLP

HCFLP is prepared as described in Example 20 by inserting a flp sequence containing EcoRI compatible ends into the EcoRI site of the lambda Zap II $V_H$ vector.

6. LCFLP

LCFLP is prepared as described in Example 20 by inserting a flp sequence containing EcoRI compatible ends into the EcoRI site of the lambda Zap II $V_L$ vector.

7. Lambda ImmunoZAP H

Lambda ImmunoZAP H is prepared by inserting the synthetic DNA sequences illustrated in Figure 25A into the above-described Lambda Zap II vector. The inserted nucleotide sequence advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper initiation of mRNA translation into protein, and a leader sequence to effieicntly direct the translated protein to the periplasm. The preparation of Lambda ImmunoZAP H is described in more detail in Example 28, and its features illustrated in Figures [25A] and [26].

8. Modified Lambda ImmunoZAP H

Modified Lambda ImmunoZAP H is prepared by inserting the modified synthetic DNA sequences illustrated in Figure 8A into the above-described Lambda ZAP II vector. The preparation of modified Lambda ImmunoZAP H and the details of the modifications are described in Example 28B. Its features are illustrated in Figure [24A] and [24B].

9. Lambda ImmunoZAP L

Lambda ImmunoZAP L is prepared as described in Example 29 by inserting into Lambda ZAP II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda ImmunoZAP L are illustrated in Figure 27.

The above-described vectors are compatible with E. coli hosts, i.e., they can express for secretion into the periplasm proteins coded for by genes to which they have been operatively linked for expression.

Examples

The following examples are intended to illustrate, but not limit, the scope of the invention.

1. Phenotype Creation

In order to obtain lambda phage clones with a range of desired phenotypes, a combinatorial library selection system was used to generate a diverse collection of clones. This approach utilized two starting populations of lambda phage clones which can be restriction digested, mixed, ligated, and packaged to form a library of clones containing DNA sequences from each of the two populations of parent phage. The following example outlines the method for rapid construction and selection of lambda phage clones containing properties from each of the two parent phage populations derived from lambda WT (cI857 ind1, Sam7) and lambda gt11 (Sam100).

Forty micrograms of a population of lambda phage derived from wild type lambda (WT) DNA (cI857 Sam7) (available from New England Biolabs) was partially digested with lambda HindIII as determined by ethidium bromide staining on 0.8% agarose gels (Maniatis et al., "Molecular Cloning," Cold Spring Harbor Laboratory (1982)). Forty micrograms of a second phage population derived from lambda gt11 DNA (available from Stratagene Cloning Systems, San Diego, CA) was digested to completion with HindIII. Subsequently, this gt11 DNA was digested

with a second enzyme BamHI in order to reduce the cloning efficiency of the left arm of the gt11 phage (Maniatis et al., supra). Both phage populations had been amplified lytically, which allowed for a relatively high degree of mutations in the resulting DNA. One microgram of the lambda WT DNA was ligated at the Hind III site to 1 to 4 μg of the lambda gt11 DNA using T4 DNA ligase in a volume of less than 20μl, according to Maniatis, et al., supra. The ligation mix was subsequently packaged in lambda phage packaging extract, Gigapack™ (Stratagene Cloning Systems, San Diego, CA), as described by the manufacturer.

The packaged phage library contained a mixture of many lambda phage constructions. In order to select for desired constructions, phenotypic selection was used to identify those members of the library displaying vigorous growth on supE bacterial hosts. As described by Maniatis et al., supra, dilutions of the phage library was plated with E. coli C600 cells (Stratagene Cloning Systems, San Diego, CA) to generate a lawn of E. coli with isolated lambda plaques. These isolated plaques are result of clonal expansion from a single lambda phage clone. Since C600 cells are supE, the growth vigor of the individual lambda phage clones could be assessed by the size of the lambda phage plaque on the E. coli lawn. The parental WT phage do not form plaques on E. coli C600. At least three classes of phage were identified and subsequently categorized as small, medium, or large plaque size. The large plaque size was an indication of vigorous growth on the phage lawn, while small plaque size indicated poor growth. This demonstrates selection for the phenotype of the S gene based on plaque size. Other pehnotypes could be used for selection.

Subsequent characterization by restriction mapping and plating on sup0 (these strains contain no amber codon suppressing tRNAs) and supF E. coli hosts, indicated that at least one of the large plaque forming clones, L2, did not contain an amber mutation as found in the lambda WT

(Sam7) or lambda gt11 (C5100) parent phage. One of the small plaque phage, S2, contained the left arm of lambda WT gene and the right arm of lambda gt11 containing the Sam100 gene. This Sam100 mutation is known to grow poorly on supE hosts and is optimal on a supF strain, with no growth on a supO host. The remaining library of clones displayed several different phenotypes, dictated by the diversity of the two starting populations of phage. Some clones also exhibited phenotypes that resulted from the random assortment of two mutant DNA fragments derived from just one of the parent DNA molecules. THis illustrates the concept that the two genes that give rise to the populations of interest need not be on separate DNA molecules at the start of the method.

Due to the phenotypic selection applied following the ligation and packaging of the phage library, the large diversity of these two populations of phage was not completely analyzed. However, the range of clones identified with alternate S gene phenotypes demonstrated some of this diversity. The diversity in these two populations of lambda phage is believed to be derived from the low level of spontaneous mutations which occur through repeated rounds of replication required in large scale preparations of lambda phage. However, the spontaneous mutations occurring within each of these individual phage populations could not generate a collection of lambda phage containing characteristics of both parent populations of phage. This combinatorial approach, therefore, provides a mechanism in which novel constructions can be generated that express genes from both parent phage constructions.

2. Polynucleotide Selection for Immunoglobulin Production

The nucleotide sequences encoding the immunoglobulin protein CDR's are highly variable. However, there are several regions of conserved sequences that flank the $V_H$ domains. For instance, contain substantially conserved nucleotide sequences, i.e., sequences that will hybridize

to the same primer sequence. Therefore, polynucleotide synthesis (amplification) primers that hybridize to the conserved sequences and incorporate restriction sites into the DNA homolog produced that are suitable for operatively linking the synthesized DNA fragments to a vector were constructed. More specifically, the DNA homologs were inserted into lambda Zap II vector (Stratagene Cloning System, San Diego, CA) at the XhoI and EcoRI sites. For amplification of the $V_H$ domains, the 3' primer (primer 67 in Table 7), was designed to be complementary to the mRNA in the $J_H$ region. In all cases, the 5' primers (primers 56-65,, Table 7) were chosen to be complementary to the first strand cDNA in the conserved N-terminus region (antisense strand). Initially amplification was performed with a mixture of 32 primers (primer 56, Table 7) that were degenerate at five positions. Hybridoma mRNA could be amplified with mixed primers, but initial attempts to amplify mRNA from spleen yielded variable results. Therefore, several alternatives to amplification using the mixed 5' primers were compared.

The first alternative was to construct multiple unique primers, eight of which are shown in Table 7, corresponding to individual members of the mixed primer pool. The individual primers 52-64 of Table 7 were constructed by incorporating either of the two possible nucleotides at three of the five degenerate positions.

The second alternative was to construct a primer containing inosine (primer 65, table 7) at four of the variable positions based on the published work of Takahashi, et al., Proc. Natl. Acad. Sci. (U.S.A.), 82:1931-1935, (1985) and Ohtsuka et al., J. Biol. Chem., 260:2605-2608, (1985). This primer has the advantage that it is not degenerate and, at the same time minimizes the negative effects of mismatches at the unconserved positions as discussed by Martin et al., Nu. Acids Res., 13:8927 (1985). However, it was not known if the presence of inosine nucleotides would result in incorporation of

unwanted sequences in the cloned $V_H$ regions. Therefore, inosine was not included at the one position that remains in the amplified fragments after the cleavage of the restriction sites. As a result, inosine was not in the cloned insert.

Additional, $V_H$ amplification primers including the unique 3' primer were designed to be complementary to a portion of the first constant region domain of the gamma 1 heavy chain mRNA (Primers 70 and 71, Table 7). These primers will produce DNA homologs containing polynucleotides coding for amino acids from the $V_H$ and the first constant region domains of the heavy chain. These DNA homologs can therefore be used to produce Fab fragments rather than an $F_v$.

As a control for amplification from spleen or hybridoma mRNA, a set of primers hybridizing to a highly conserved region within the constant region IgG, heavy chain gene were constructed. The 5' primer (primer 66, Table 7) is complementary to the cDNA in the $C_H2$ region whereas the 3' primer (primer 68, Table 7) is complementary to the mRNA in the $C_H3$ region. It is believed that no mismatches were present between these primers and their templates.

The nucleotide sequences encoding the $V_L$ CDRs are highly variable. However, there are several regions of conserved sequences that flank the $V_L$ CDR domains including the $J_L$, $V_L$ framework regions and $V_L$ leader/promoter. Therefore, amplification primers that hybridize to the conserved sequences and incorporate restriction sites that allowing cloning the amplified fragments into the pBluescript SK-vector cut with Nco I and SpeI were constructed. For amplification of the $V_L$ CDR domains, the 3' primer (primer 69 in Table 7), was designed to be complementary to the mRNA in the $J_L$ regions. The 5' primer (primer 70, Table 7) was chosen to be complementary to the first strand cDNA in the conserved N-terminus region (antisense strand).

A second set of amplification primers for amplification of the $V_L$ CDR domains the 5' primers (primers 73-80 in Table 8) were designed to be complementary to the first strand cDNA in the conserved N-terminus region. These primers also introduced a Sac I restriction endonuclease site to allow the FLDNA homolog to be cloned into the $V_L$II-expression vector. The 3' $V_L$ amplification primer (primer 81 in Table 8) was designed to be complementary to the mRNA in the $J_L$ regions and to introduce the XbaI restriction endonuclease site required to insert the $V_L$DNA homolog into the $V_L$II-expression vector (Figure 8).

Additional 3' $V_L$ amplification primers were designed to hybridize to the constant region of either kappa or lambda mRNA (primers 82 and 83 in Table 8). These primers allow a DNA homolog to be produced containing polynucleotide sequences coding for constant region amino acids of either kappa or lambda chain. These primers make it possible to produce an Fab fragment rather than an $F_v$.

The primers used for amplification of kappa light chain sequences for construction of Fabs are shown at least in Table 8. Amplification with these primers was performed in 5 separate reactions, each containing one of the 5' primers (primers 75-78, and 84) and one of the 3' primer (primer 81) has been used to construct $F_v$ fragments. The 5' primers contain a Sac I restriction site and the 3' primers contain a XbaI restriction site.

The primers used for amplification of heavy chain Fd fragments for construction of Fabs are shown at least in Table 7. Amplification was performed in eight separate reactions, each containing one of the 5' primers (primers 57-64) and one of the 3' primers (primer 70). The remaining 5' primers that have been used for amplification in a single reaction are either a degenerate primer (primer 56) or a primer that incorporates inosine at four degenerate positions (primer 66, Table 7, and primers 89 and 90, Table 8). The remaining 3' primer (primer 86, Table 8) has been used to construct $F_v$ fragments. Many of the 5'

primers incorporate a Xho I site, and 3' primers include a SpeI restriction site.

$V_L$ amplification primers designed to amplify human light chain variable regions of both the lambda and kappa isotypes are also shown in Table 8.

All primers and synthetic polynucleotides used herein and shown on Tables 7-11 were either purchased from Research Genetics in Huntsville, Alabama or synthesized on an Applied Biosystems DNA synthesizer, model 381A, using the manufacturer's instructions.

TABLE 7

| | |
|---|---|
| (56) 5' AGGT(C/G)(C/A)A(G/A)CT(G/T)CTCGAGTC(T/A)GG 3' | degenerate 5' primer for the amplification of variable heavy chain region ($V_H$) |
| (57) 5' AGGTCCAGCTGCTCGAGTCTGG 3' | Unique 5' primer for the amplification of $V_H$ |
| (58) 5' AGGTCCAGCTGCTCGAGTCAGG 3' | " |
| (59) 5' AGGTCCAGCTTCTCGAGTCTGG 3' | " |
| (60) 5' AGGTCCAGCTTCTCGAGTCAGG 3' | " |
| (61) 5' AGGTCCAACTGCTCGAGTCTGG 3' | " |
| (62) 5' AGGTCCAACTGCTCGAGTCAGG 3' | " |
| (63) 5' AGGTCCAACTTCTCGAGTCTGG 3' | " |
| (64) 5' AGGTCCAACTTCTCGAGTCAGG 3' | " |
| (65) 5' AGGTIIAICTICTCGAGTC(T/A) 3' | 5' degenerate primer containing inosine at 4 degenerate positions |
| (66) 5' GCCCAAGGATGTGCTCACC 3' | 5' primer for amplification in the $C_H2$ region of mouse IgG1 |
| (67) 5' CTATTAGAATTCAACGGTAACAGTGGTGCCTTGGCCCCA 3' | 3' primer for amplification of $V_H$ |
| (67A) 5' CTATTAACTAGTAACGGTAACAGTGGTGCCTTGGCCCCA 3' | 3' primer for amplification of $V_H$ using 3' spe I site |
| (68) 5' CTCAGTATGGTGGTTGTGC 3' | 3' primer for amplification in the $C_H3$ region of mouse IgG1 |
| (69) 5' GCTACTAGTTTTGATTTCCACCTTGG 3' | 3' primer for amplification of $V_L$ |

(70) 5' CAGCCATGGCCGACATCCAGATG 3' 5' primer for amplification of $V_L$

(71) 5' AATTTTACTAGTCACCTTGGTGCTGCTGGC 3' Unique 3' primer for amplification of $V_H$ including part of the mouse gamma 1 first constant

(72) 5' TATGCAACTAGTACAACCACAATCCCTGGGCACAATTTT 3' Unique 3' primer for amplification of $V_H$ including part of mouse gamma 1 first constant region and hinge region

TABLE 8

| | | |
|---|---|---|
| (73) | 5' CCAGTTCCGAGCTCGTTGTGACTCAGGAATCT 3' | Unique 5' primer for the amplification of $V_L$ |
| (74) | 5' CCAGTTCCGAGCTCGTGTTGACGCAGCCGCCC 3' | " |
| (75) | 5' CCAGTTCCGAGCTCGTGCTCACCCAGTCTCCA 3' | " |
| (76) | 5' CCAGTTCCGAGCTCCAGATGACCCAGTCTCCA 3' | " |
| (77) | 5' CCAGATGTGAGCTCGTGATGACCCAGACTCCA 3' | " |
| (78) | 5' CCAGATGTGAGCTCGTCATGACCCAGTCTCCA 3' | " |
| (79) | 5' CCAGATGTGAGCTCTTGATGACCCAAACTCAA 3' | " |
| (80) | 5' CCAGATGTGAGCTCGTGATAACCCAGGATGAA 3' | " |
| (81) | 5' GCAGCATTCTAGAGTTTCAGCTCCAGCTTGCC 3' | Unique 3' primer for $V_L$ amplification |
| (82) | 5' CCGCCGTCTAGAACACTCATTCCTGTTGAAGCT 3' | Unique 3' primer for $V_L$ amplification including the kappa constant region |
| (83) | 5' CCGCCGTCTAGAACATTCTGCAGGAGACAGACT 3' | Unique 3' primer for $V_L$ amplification including the lambda constant region |
| (84) | 5' CCAGTTCCGAGCTCGTGATGACACAGTCTCCA 3' | Unique 5' primer for $V_L$ amplification |
| (85) | 5' GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA 3' | Unique 3' primer for $V_L$ amplification |
| (86) | 5' CTATTAACTAGTAACGGTAACAGTGGTGCCTTGCCCCA 3' | " |
| (87) | 5' AGGCTTACTAGTACAATCCCTGGGCACAAT 3' | Unique 3' primer for $V_H$ amplification |
| (88) | 5' GCCGCTCTAGAACACTCATTCCTGTTGAA 3' | Unique 3' primer for $V_L$ amplification |
| (89) | 5' AGGTIIAICTICTCGAGTCTGC 3' | Degenerate 5' primer containing inosine at 4 degenerate positions |
| (90) | 5' AGGTIIAICTICTCGAGTCAGC 3' | " |

3. Production Of A $V_H$ Coding Repertoire Enriched In FITC Binding Proteins

Fluorescein isothiocyanate (FITC) was selected as a ligand for receptor binding. It was further decided to enrich by immunization the immunological gene repertoire, i.e., $V_H$- and $V_L$-coding gene repertoires, for genes coding for anti-FITC receptors. This was accomplished by linking FITC to keyhole limpet hemocyanin (KLH) using the techniques described in Antibodies A Laboratory Manual, Harlow and Lowe, eds., Cold Spring Harbor, New York, (1988). Briefly, 10.0 milligrams (mg) of keyhole limpet hemocyanin and 0.5 mg of FITC were added to 1 ml of buffer containing 0.1 M sodium carbonate at pH 9.6 and stirred for 18 to 24 hours at 4 degrees C (4C). The unbound FITC was removed by gel filtration through Sephadex G-25.

The KLH-FITC conjugate was prepared for injunction into mice by adding 100 μg of the conjugate to 250 μl of phosphate buffered saline (PBS). An equal volume of complete Freund's adjuvant was added and the entire solution was emulsified for 5 minutes. A 129 $G_{IX+}$ mouse was injected with 300 μl of the emulsion. Injections were given subcutaneously at several sites using a 21 gauge needle. A second immunization with KLH-FITC was given two week later. This injection was prepared as follows: fifty μg of KLH-FITC were diluted in 250 μL of PBS and an equal volume of alum was admixed to the KLH-FITC solution. The mouse was injected intraperitoneally with 500 μl of the solution using a 23 gauge needle. One month later the mice were given a final injection of 50 _g of the KLH-FITC conjugate diluted to 200 _L in PBS. This injection was given intravenously in the lateral tail vein using a 30 gauge needle. Five days after this final injection the mice were sacrificed and total cellular RNA was isolated from their spleens.

Hybridoma PCP 8D11 producing an antibody immunospecific for phosphonate ester was cultured in DMEM media (Gibco Laboratories, Grand Island, New York) containing 10

percent fetal calf serum supplemented with penicillin and streptomycin. About 5 x $10^8$ hybridoma cells were harvested and washed twice in phosphate buffered saline. Total cellular RNA was prepared from these isolated hybridoma cells.

4. Preparation Of A $V_H$-Coding Gene Repertoire

Total cellular RNA was prepared from the spleen of a single mouse immunized with KLH-FITC as described in Example 3 using the RNA preparation methods described by Chomczynski et al., Anal Biochem., 162:156-159 (1987) using the manufacturer's instructions and the RNA isolation kit produced by Stratagene Cloning Systems, La Jolla, CA. Briefly, immediately after removing the spleen from the immunized mouse, the tissue was homogenized in 10 ml of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M 2-mercaptoethanol using a glass homogenizer. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized spleen. One ml of phenol that had been previously saturated with $H_2O$ was also admixed to the denaturing solution containing the homogenized spleen. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuge tube (Fisher Scientific Company, Pittsburgh, PA). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described

above. Three ml of isopropyl alcohol was added to the resuspended total cellular RNA and vigorously mixed. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for 15 minutes and then resuspended to dimethyl pyrocarbonate (DEPC) treated $H_2O$ (DEPC-$H_2O$).

Messenger RNA (mRNA) enriched for sequences containing long poly A tracts was prepared from the total cellular RNA using methods described in Molecular Cloning A Laboratory Manual, Maniatias et al., eds. Cold Spring Harbor Laboratory, New York, (1982). Briefly, one half of the total RNA isolated from a single immunized mouse spleen prepared as described above was resuspended in one ml of DEPC-$H_2O$ and maintained at 65°C for five minutes. One ml of 2x high salt loading buffer consisting of 100 mM Tris-HCl, 1M sodium chloride, 2.0 mM disodium ethylene diamine tetraacetic acid (EDTA) at pH 7.5, and 0.2% sodium dodecyl sulfate (SDS) was added to the resuspended RNA and the mixture allowed to cool to room temperature. The mixture was then applied to an oligo-dT (Collaborative Research Type 2 or Type 3) column that was previously prepared by washing the oligo-dT with a solution containing 0.1 M sodium hydroxide and 5 mM EDTA and then equilibrating the column with DEPC-$H_2O$. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for 5 minutes at 65°C. The oligo dT column was then washed with 2 ml of high salt loading buffer consisting of 50 mM Tris-HCl at pH 7.5, 500 mM sodium chloride, 1 mM EDTA at pH 7.5 and 0.1% SDS. The oligo dT column was then washed with 2 ml of 1 X medium salt buffer consisting of 50 mM Tris-HCl at pH 7.5, 100 mM sodium chloride 1 mM EDTA and 0.1% SDS. The messenger RNA was eluted from the oligo dT column with 1ml of buffer consisting of 10 mM Tris-HCL at pH 7.5, 1 mM

EDTA at pH 7.5 and 0.05% SDS. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and resuspended in DEPC- $H_2O$.

The messenger RNA isolated by the above process contains a plurality of different $V_H$ coding polynucleotides, i.e., greater than about $10^4$ different $V_H$-coding genes.

5. Preparation Of A Single $V_H$ Coding Polynucleotide

Polynucleotides coding for a single $V_H$ were isolated according to Example 4 except total cellular RNA was extracted from monoclonal hybridoma cells prepared in Example 3. The polynucleotides isolated in this manner code for a single $V_H$.

6. DNA Homolog Preparation

In preparation for PCR amplification, mRNA prepared according to the above examples was used as a template for cDNA synthesis by a primer extension reaction. In a typical 50 ul transcription reaction, 5-10 ug of spleen or hybridoma mRNA in water was first hybridized (annealed) with 500 ng (50.0 pmol) of the 3' $V_H$ primer (primer 67, Table 7), at 65°C for five minutes. Subsequently, the mixture was adjusted to 1.5 mM dATP, dCTP, dGTP and dTTP, 40 mM Tris-HCl at pH 8.0, 8 mM $MgCl_2$, 50 mM NaCl, and 2 mM spermidine. Moloney-Murine Leukemia virus Reverse transcriptase (Stratagene Cloning Systems), 26 units, was added and the solution was maintained for 1 hours at 37°C.

PCR amplification was performed in a 100 ul reaction containing the products of the reverse transcription reaction (approximately 5 ug of the cDNA/RNA hybrid), 300 ng of 3' $V_H$ primer (primer 67 of Table 7), 300 ng each of the 5' $V_H$ primers (primer 57-65 of Table 7) 200 mM of a mixture of dNTP's, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$, 0.1% gelatin and 2 units of Taq DNA polymerase. The

reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92°C for 1 minute, annealing at 52°C for 2 minutes and polynucleotide synthesis by Primer extension (elongation) at 72°C for 1.5 minutes. The amplified $V_H$-coding DNA homolog containing samples were extracted twice with phenol/chloroform, once with chloroform, ethanol precipitated and were stored at -70°C in 10 mM Tris-HCl, (pH, 7.5) and 1 mM EDTA.

Using unique 5' primers (57-64, Table 7), efficient $V_H$-coding DNA homolog synthesis and amplification from the spleen mRNA was achieved as shown in Figure 3, lanes R17-R24. The amplified cDNA ($V_H$-coding DNA homolog) is seen as a major band of the expected size (360 bp). The intensities of the amplified $V_H$-coding polynucleotide fragment in each reaction appear to be similar, indicating that all of these primers are about equally efficient in initiating amplification. The yield and quality of the amplification with these primers was reproducible.

The primer containing inosine also synthesized amplified $V_H$-coding DNA homologs from spleen mRNA reproducibly, leading to the production of the expected sized fragment, of an intensity similar to that of the other amplified cDNAs (Figure 4, lane R16). This result indicated that the presence of inosine also permits efficient DNA homolog synthesis and amplification. Clearly indicating how useful such primers are in generating a plurality of $V_H$-coding DNA homologs. Amplification products obtained from the constant region primers (primers 66 and 68, Table 7) were more intense indicating that amplification was more efficient, possibly because of a higher degree of homology between the template and primers (Figure 4, Lang R9). Based on these results, a $V_H$-coding gene library was constructed from the products of eight amplifications, each performed with a different 5' primer. Equal portions of the products from each primer extension reaction were

mixed and the mixed product was then used to generate a library of $V_H$-coding DNA homolog-containing vectors.

DNA homologs of the $V_L$ were prepared from the purified mRNA prepared as described above. In preparation for PCR amplification, mRNA prepared according to the above examples was used as a template for cDNA synthesis. In a typical 50 ul transcription reaction, 5-10 ug of spleen or hybridoma mRNA in water was first annealed with 300 ng (50.0 pmol) of the 3' $V_L$ primer (primer 69, Table 7), at 65°C for five minutes. Subsequently, the mixture was adjusted to 1.5 mM dATP, dCTP, dGTP, and dTTP, 40 mM Tris-HCL at pH 8.0, 8 mM $MgCl_2$, 50 mM NaCl, and 2 mM spermidine. Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), 26 units, was added and the solution was maintained for 1 hour at 37°C. The PCR amplification was performed in a 100 ul reaction containing approximately 5 ug of the cDNA/RNA hybrid produced as described above, 300 ng of the 3' $V_L$ primer (primer 69 of Table 7), 300 ng of the 5' $V_L$ primer (primer 70 of Table 7), 200 mM of a mixture of dNTP's, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$, 0.1% gelatin and 2 units of Taq DNA polymerase. The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle involved denaturation at 92°C for 1 minute, annealing at 52°C for 2 minutes and elongation at 72°C for 1.5 minutes. The amplified samples were extracted twice with phenol/chloroform, once with chloroform, ethanol precipitated and were stored at 70°C in 10 mM Tris-HCl at pH 7.5 and 1 mM EDTA.

7. Inserting DNA Homologs Into Vectors

In preparation for cloning a library enriched in $V_H$ sequences, PCR amplified products (2.5 mg/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 MM $MgSo_4$, 1 mM DTT, 200 mg/ml bovine serum albumin (BSA) at 37°C were digested with restriction enzymes Xho I (125 units) and EcoR I (10 U) and purified on a 1% agarose gel. In cloning experi-

ments which required a mixture of the products of the amplification reactions, equal volumes (50 ul, 1-10 ug concentration) of each reaction mixture were combined after amplification but before restriction digestion. After gel electrophoresis of the digested PCR amplified spleen mRNA, the region of the gel containing DNA fragments of approximately 350 bps was excised, electroeluted into a dialysis membrane, ethanol precipitated and resuspended in 10 mM Tris-HCl pH 7.5 and 1 mM EDTA to a final concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 5°C to 1 ug of Lambda Zap™ II vector (Stratagene Cloning Systems, La Jolla, CA) previously cut by EcoR I and Xho I. A portion of the ligation mixture (1 ul) was packaged for 2 hours at room temperature using Gigapack Gold packaging extract (Stratagene Cloning Systems, La Jolla, CA), and the packaged material was plated on IL1-blue host cells. The Library was determined to consist of 2 x $10^7$ $V_H$ homologs with less than 30% non-recombinant background.

The vector used above, Lambda Zap II is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the form of a phagemid (Bluescript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, (1988), by replacing the Lambda S gene contained in a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard proto-

cols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, New York, (1987).

In preparation of cloning a library enriched in $V_L$ sequences, 2 ug of PCR amplified products (2.5 mg/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM $MgSo_4$, 1 mM DTT, 200 mg/ml BSA) were digested with restriction enzymes Nco I (30 unites) and Spe I (45 units) at 37°C for 2 hours. The digested PCR amplified products were purified on 1% agarose gel using standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). Briefly, after gel electroelution of the digested PCR amplified product the region of the gel containing the $V_L$-coding DNA fragment of the appropriate size was excised, electroelution into a dialysis membrane, ethanol precipitated and resuspended at a final concentration of 10 ng per ml in a solution containing 10 mM Tris-HCL at pH 7.5 and 1 mM EDTA.

An equal molar amount of DNA representing a plurality of different $V_L$-coding DNA homologs was ligated to a pBluescript SK- phagemid vector that had been previously cut with Nco I and Spe I. A portion of the ligation mixture was transformed using the manufacturer's instructions into Epicuian Coli XL1-Blue competent cells (Stratagene Cloning Systems, La Jolla, CA). The transformant library was determined to consist of $1.2 \times 10^3$ colony forming units/ug of $V_L$ homologs with less than 3% non-recombinant background.

8. Sequencing of Plasmids From the $V_H$-Coding cDNA Library

To analyze the Lambda Zap II phage clones, the clones were excised from Lambda Zap into plasmids according to the manufacture's instructions (Stratagene Cloning System, La Jolla, CA). Briefly, phage plaques were cored from the agar plates and transferred to sterile microfuge tubes containing 500 ul a buffer containing 50 mM Tris-HCL at pH

7.5, 100 mM NaCl, 10 mM $MgSO_4$, and 0.01% gelatin and 20 uL of Chloroform.

For excisions, 200 ul of the phage stock, 200 ul of XL1-Blue cells ($A_{600}$ = 1.00) and 1 ul of R408 helper phage ($1 \times 10^{11}$ pfu/ml) were incubated at 37°C for 15 minutes. The excised plasmids were infected into XL1-Blue cells and plated onto LB plates containing ampicillin. Double stranded DNA was prepared from the phagemid containing cells according to the methods described by Holmes et al., Anal. Biochem., 114:193, (1981). Clones were first screened for DNA inserts by restriction digests with either Pvu II or Bg1 I and clones containing the putative $V_H$ insert were sequenced using reverse transcriptase according to the general method described by Sanger et al., Proc. Natl. Acad. Sci., USA, 74:5463-5467, (1977) and the specific modifications of this method provided in the manufacturer's instruction in the AMV reverse transcriptase $^{35}$S-dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA.

9. Characterization Of The Cloned $V_H$ Repertoire

The amplified products which had been digested with Xho I and EcoR I and cloned into Lambda Zap, resulted in a cDNA library with $9.0 \times 10^5$ pfu's. In order to confirm that the library consisted of a diverse population of $V_H$-coding DNA homologs, the N-terminal 120 bases of 18 clones, selected at random from the library, were excised and sequenced (Figure 5). To determine if the clones were of $V_H$ gene origin, the cloned sequences were compared with known $V_H$ sequences and $V_L$ sequences. The clones exhibited from 80 to 90% homology with sequences of known heavy chain origin and little homology with sequences of light chain origin when compared with the sequences available in Sequences of Proteins of Immunological Interest by Kabot et al., 4th ed., U.S. Dept. of Health and Human Sciences, (1987). This demonstrated that the library was enriched

for the desired $V_H$ sequence in preference to other sequences, such as light chain sequences.

The diversity of the population was assessed by classifying the sequenced clones into predefined subgroups (Figure 5). Mouse $V_H$ sequences are classified into eleven subgroups (Figure 5). Mouse $V_H$ sequences are classified into eleven subgroups [I (A,B,), II (A,B,C), III (A,B,C,D( V (A,B)] based on framework amino acid sequences described in Sequences of proteins of Immunological Interest by Kabot et al., 4th ed., U.S. Dept. of Health and Human Sciences, (1987); Dildrop, Immunology Today, 5:84, (1984); and Brodeur et al., Eur. J. Immunol., 14:922, (1984). Classification of the sequenced clones demonstrated that the cDNA library contained $V_H$ sequences of at least 7 different subgroups. Further, a pairwise comparison of the homology between the sequenced clones showed that no two sequences were identical at all positions, suggesting that the population is diverse to the extent that it is possible to characterize by sequence analysis.

Six of the clones (L 36-50, Figure 5) belong to the subclass III B and had very similar nucleotide sequences. This may reflect a preponderance of mRNA derived from one or several related variable genes in stimulated spleen, but the data does not permit ruling out the possibility of a bias in the amplification process.

10. $V_H$-Expression Vector Construction

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is a highly efficient method of reintroducing DNA into host cells. Second, it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. An alternative,

plasmid cloning vectors, are only advantageous in the analysis of clones after they have been identified. This advantage is not lost in the present system because of the use of lambda Zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be excised.

To express the plurality of $V_H$-coding DNA homologs in an E. coli host cell, a vector was constructed that placed the $V_H$-coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the $V_H$-coding DNA homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6. The individual single-stranded polynucleotides (N1-N12) are shown in Table 9.

Polynucleotides N2, N3, N9-4', N11, N10-5', N6, N7 and N8 were kinased by adding 1 ul of each polynucleotide (0.1 ug/ul) and 20 units of T4 polynucleotide kinase to a solution containing 70 mM Tris-HCl at pH 7.6, 10 mM $MgCl_2$ 5 mM DTT, 10 mM 2-mercaptoethanol (2ME), 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCl at pH 7.4, 2.0 mM $MgCl_2$ and 50.0 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in

a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 6A. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 ul of the above reaction to a solution containing 50 mM Tris-HCl at pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM adenosine triphosphate (ATP) and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 ul of the above reaction, 4 ul of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1 blue cells (Stratagene Cloning Systems, San Diego, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxide method described in by Sanger et al., Proc. Natl. Acad. Sci USA, 74:5463-5467, (1977) and using the manufacture's instruction in the AMV Reverse Transcriptase $^{35}S$-ATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting $V_H$ expression vector is shown in Figure 6A and Figure 7.

Table 9

(91) N1) 5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
(92) N2) 5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
(93) N3) 5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'
(94) N4) 5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTTAATAG 3'
(95) N5) 5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
(96) N6) 5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3'
(97) N7) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(98) N8) 5' GTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGC 3'
(99) N9-4) 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
(100) N11) 5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
(101) N12) 5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
(102) N10-5) 5 CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

11. $V_L$ Expression Vector Construction

To express the plurality of $V_L$ coding polynucleotides in an E. coli host cell, a vector was constructed that placed the $V_L$ coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the piroplasmic space and also provided a polynucleotide that coded for a spacer protein between the $V_L$ polynucleotide and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6B. The individual single-stranded polynucleotides (N1-N8) are shown in Table 9.

Polynucleotides N2, N3, N4, N6, N7 and N8 were kinased by adding 1 ul of each polynucleotide and 20 units of T4 polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintain-

ing the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCL at pH 7.4, 2.0 mM $MgCl_2$ and 50.0 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40 ul of the above reaction to a solution containing 50 ul Tris-HCL at pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 10 units to T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 ul of the above reaction, 4 ul of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1-Blue cells (Stratagene Cloning Systems, La Jolla, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA and described in Short et al., Nucleic Acids Res., 16:7583-7600 (1988). This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a phagemid vector to

allow easy manipulation and sequencing and also produces the phagemid version of the $V_L$ expression vectors. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxide method described by Sanger et al., Proc. Natl. Acad. Aci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase $^{35}$S-dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting $V_L$ expression vector is shown in Figure 6 and Figure 8.

The $V_L$ expression vector used to construct the $V_L$ library was the phagemid produced to allow the DNA of the $V_L$ expression vector to be determined. The phagemid was produced, as detailed above, by the in vivo excision process from the Lambda Zap $V_L$ expression vector (Figure 8). The phagemid version of this vector was used because the Nco I restriction enzyme site is unique in this version and thus could be used to operatively linked the $V_L$ DNA homologs into the expression vector.

12. $V_L$II-Expression Vector Construction

To express the plurality of $V_L$-coding DNA homologs in an E. coli host cell, a vector was constructed that placed the $V_L$-coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided the Pel B gene leader sequence that has been previously used to successfully secrete Fab fragments in E. coli by Lei et al., J. Bac., 169:4379 (1987) and Better et al., Science, 240:1041 (1988), and also provided a polynucleotide containing a restriction endonuclease site for cloning. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-60 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 10. The sequence of each individual single-stranded polynucleotides (01-08) within

the double stranded synthetic DNA sequence is shown in Table 10.

Polynucleotides 02, 03, 04, 05, 06 and 07 were kinased by adding 1 ul (0.1 ug/ul) of each polynucleotide and 20 units of T4 polynucleotide kinase to a solution containing 70 mM Tris-HCL at pH 7.6, 10 mM magnesium chloride (MgCl), 5 mM dithiothreitol (DTT), 10 mM 2-mercaptoethanol (2ME), 500 micrograms per ml of bovine serum albumin. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The 20 ng each of the two end polynucleotides, 01 and 08, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20.0 mM Tris-HCl at pH 7.4, 2.0 mM MgCl and 15.0 mM sodium chloride (NaCl). This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all 8 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 9. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 ul of the above reaction to a solution containing 50 ml Tris-HCl at pH 7.5, 7 ml MgCl, 1 mm DTT, 1 mm ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 ul of the above reaction, 4 ul of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes. The completed synthetic DNA insert was ligated directly into a lambda Zap II vector that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacturer's instructions

using Gigapack II Gold packing extract available from Stratagene Cloning Systems, La Jolla, CA. The packaged ligation mixture was plated on XL1 blue cells (Stratagene Cloning Systems, San Diego, CA). Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer, Stratagene Cloning Systems, La Jolla, CA. This in vivo excision protocol moves the cloned insert from the lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the manufacturer's instructions in the AMV Reverse Transcriptase $^{35}$S-dATP sequencing kit from Stratagene Cloning Systems, La Jolla, CA. The sequence of the resulting $V_L$II-expression vector is shown in Figure 9 and Figure 11.

Table 10

(102) 01) 5' TGAATTCTAAACTAGTCGCCAAGGAGACAGTCAT 3'
(103) 02) 5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
(104) 03) 5' GTTATTACTCGCTGCCCAACCAGCCATGGCC 3'
(105) 04) 5' GAGCTCGTCAGTTCTAGAGTTAAGCGGCCG 3'
(106) 05) 5' GTATTTCATTATGACTGTCTCCTTGGCGACTAGTTTAGAATTCAAGCT 3'
(107) 06) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(108) 07) 5' TGACGAGCTCGGCCATGGCTGGTTGGG 3'
(109) 08) 5' TCGACGGCCGCTTAACTCTAGAAC 3'

13. $V_H + V_L$ Library Construction

To prepare an expression library enriched in $V_H$ sequences, DNA homologs enriched in $V_H$ sequences were prepared according to Example 7 using the same set of 5' primers but with primer 62A (Table 7) as the 3' primer. These homologs were then digested with the restriction enzymes Xho I and Spe I and purified on a 1% agarose gel using the standard electroelution technique described in

Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). These prepared $V_H$ DNA homologs were then directly inserted into the $V_H$ expression vector that had been previously digested with Xho I and Spe I.

The ligation mixture containing the $V_H$ DNA homologs were packaged according to the manufacturers specifications using Gigapack Gold II Packing Extract (Stratagene Cloning Systems, La Jolla, CA). The expression libraries were then ready to be plated on XL-1 Blue cells.

To prepare a library enriched in $V_L$ sequences, PCR amplified products enriched in $V_L$ sequences were prepared according to Example 7. The $V_L$ DNA homologs were digested with restriction enzymes Nco I and Spe I. The digested $V_L$ DNA homologs were purified on a 1% agarose gel using standard electrelusion techniques described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY (1982). The prepared $V_L$ DNA homologs were directly inserted into the $V_L$ expression vector that had been previously digested with the restriction enzymes Nco I and Spe I. The ligation mixture containing the $V_L$ DNA homologs were transformed into XL-1 blue competent cells using the manufacturer's instructions (Stratagene Cloning Systems, La Jolla, CA).

14. Inserting $V_L$ Coding DNA Homologs Into $V_L$ Expression Vector

In preparation for cloning a library enriched in $V_L$ sequences, PCR amplified products (2.5 ug/30 ul of 150 mM NaCl, 8 mM Tris-HCl (pH 7.5), 6 mM $MgCl_4$, 1 mM DTT, 200 ug/ml BSA at 37°C were digested with restriction enzymes Sac I (125 units) and Xba I (125 units) and purified on a 1% agarose gel. In cloning experiments which required a mixture of the products of the amplification reactions, equal volumes (50 ul, 1-10 ug concentration) of each reaction mixture were combined after amplification but before restriction digestion. After gel electrophoresis

of the digested PCR amplified spleen mRNA, the region of the gel containing DNA fragments of approximately 350 bps was excised, electroeluted into a dialysis membrane, ethanol precipitated and resuspended in a TE solution containing 10 mM Tris-HCl pH 7.5 and 1 mM EDTA to a final concentration of 50 ng/ul.

The $V_L$II-expression DNA vector was prepared for cloning by admixing 100 ug of this DNA to a solution containing 250 units each of the restriction endonucleases Sac 1 and Xba 1 (both from Boehringer Mannheim, Indianapolis, IN) and a buffer recommended by the manufacturer. This solution was maintained at 37°C for 1.5 hours. The solution was heated at 65°C for 15 minutes to inactivate the restriction endonucleases. The solution was chilled to 30°C and 25 units of heat-killable (HK) phosphatase (Epicenter, Madison, WI) and $CaCl_2$ were admixed to it according to the manufacturer's specifications. This solution was maintained at 30°C for 1 hour. The DNA was purified by extracting the solution with a mixture of phenol and chloroform followed by ethanol precipitation. The $V_L$II expression vector was now ready for ligation to the $V_L$ DNA homologs prepared in the above examples.

DNA homolog enriched in $V_L$ sequences were prepared according to Example 6 but using a 5' light chain primer and 3' light chain primer shown in Table 9. Individual amplification reactions were carried out using each 5' light chain primer in combination with the 3' light chain primer. These separate $V_L$ homolog-containing reaction mixtures were mixed and digested with the restriction endonucleases Sac 1 and Xba 1 according to Example 7. The $V_L$ homologs were purified on a 1% agarose gel using the standard electroelution technique described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, New York, (1982). These prepared $V_L$ DNA homologs were then directly inserted into the Sac 1 - Xba cleaved $V_L$II-expression vector that was prepared above by ligating 3 moles of $V_L$ DNA homolog inserts with each mole

of the $V_L$II-expression vector overnight at 5°C. $3.0 \times 10^5$ plaque forming units were obtained after packaging the DNA with Gigapack II Bold (Stratagene Cloning Systems, La Jolla, CA) and 50% were recombinants.

15. Randomly Combining $V_H$ and $V_L$ DNA Homologs on the Same Expression Vector

The $V_L$II-expression library prepared in Example 13 was amplified and 500 ug of $V_L$II-expression library phage DNA prepared from the amplified phage stock using the procedures described in Molecular Cloning: A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982), 50 ug of this $V_L$II-expression library phage DNA was maintained in a solution containing 100 units of LuI restriction endonuclease (Boehringer Mannheim, Indianapolis, IN) in 200 ul of a buffer supplied by the endonuclease manufacturer for 1.5 hours at 37°C. The solution was then extracted with a mixture of phenol and chloroform. The DNA was then ethanol precipitated and resuspended in 100 ul of water. This solution was admixed with 100 units of the restriction endonuclease EcoR I (Boehringer Mannheim, Indianapolis, IN) in a final volume of 200 ul of buffer containing the components specified by the manufacturer. This solution was maintained at 37°C for 1.5 hours and the solution was then extracted with a mixture of phenol and chloroform. The DNA was ethanol precipitated and the DNA resuspended in TE.

The $V_H$ expression library prepared in Example 13 was amplified and 500 ug of $V_H$ expression library phage DNA prepared using the methods detailed above. 50 ug of the $V_H$ expression library phage DNA was maintained in a solution containing 100 units of Hind III restriction endonuclease (Boehringer Mannheim, Indianapolis, IN) in 200 ul of a buffer supplied by the endonuclease manufacturer for 1.5 hours at 37°C. The solution was then extracted with a mixture of phenol and chloroform saturated with 0.1

Tris-HCL at pH 7.5. The DNA was then ethanol precipitated and resuspended in 100 ul of water. This solution was admixed with 100 units of the restriction endonuclease EcoR I (Boehringer Mannheim, Indianapolis, IN) in a final volume of 200 ul of buffer containing the components specified by the manufacturer. This solution was maintained at 37°C for 1.5 hours and the solution was then extracted with a mixture of phenol and chloroform. The DNA was ethanol precipitated and the DNA resuspended in TE.

The restriction digested $V_H$ and $V_L$II-expression Libraries were ligated together. The ligation reaction consisted of 1 ug of $V_H$ and 1 ug of $V_L$II phage library DNA in a 10 ul reaction using the reagents supplied in a ligation kit purchased from Stratagene Cloning Systems (La Jolla, CA). After ligation for 16 hr at 4°C, 1 ul of the ligated the phage DNA was packaged with Gigapack Gold II packaging extract and plated on XL 1-blue cells prepared according the manufacturer's instructions. A portion of the $3X10^6$ clones obtained were used to determine the effectiveness of the combination. The resulting $V_H$ and $V_L$ expression vector is shown in Figure 11.

Clones containing both $V_H$ and $V_L$ were excised from the phage to pBluescript using the in vitro excision protocol described by Short et al., Nucleic Acid Research, 16L7583-7600 (1988). Clones chosen for excision expressed the decapeptide tag and did not cleave X-gal in the presence of 2mM IPTG, thus remaining white. Clones with these characteristics represented 30% of the library. 50% of the clones chosen for excision contained a $V_H$ and $V_L$ as determined by restriction analysis. Since approximately 30% of the clones in the $V_H$ library expressed the decapeptide tag and 50% of the clones in the $V_L$II library contained a $V_L$ sequence it was anticipated that no more than 15% of the clones in the combined library would contain both $V_H$ and $V_L$ clones. The actual number obtained

was 15% of the library indicating that the process of combination was very efficient.

16. Segregating DNA Homologs For a $V_H$ Antigen Binding Protein

To segregate the individual clones containing DNA homologs that code for a $V_H$ antigen binding protein, the titre of the $V_H$ expression library prepared according to Example 12 was determined. This library titration was performed using methods well known to one skilled in the art. Briefly, serial dilutions of the library were made into a buffer containing 100 mM NaCl, 50 mM Tris-HCL at pH 7.5 and 10 mM $MgCl_4$, 5 g/L yeast extract, 10 g/L NZ amine (casein hydrolysate) and 0.7% melted, 50°C agarose. The phage, the bacteria and the top agar were mixed and then evenly distributed across the surface of a prewarmed bacterial agar plate (5 g/L NaCl, 2 g/L $MgCl_4$, 5 g/L yeast extract, 10 g/L NZ amine (casein hydrolysate) and 15 g/L Difco agar. The plates were maintained at 37°C for 12 to 24 hours during which time period the lambda plaques developed on the bacterial lawn. The lambda plaques were counted to determine the total number of plaque forming units per ml in the original library.

The titred expression library was then plated out so that replica filters could be made from the library. The replica filters will be used to later segregate out the individual clones in the library that are expressing the antigens binding proteins of interest. Briefly, a volume of the titred library that would yield 20,000 plaques per 150 millimeter plate was added to 600 ul of exponentially growing E. coli cells and maintained at 37°C for 15 minutes to allow the phage to absorb to the bacterial cells. The 7.5 ml of top agar was admixed to the solution containing the bacterial cells and the absorbed phage and the entire mixture distributed evenly across the surface of a prewarmed bacterial agar plate. This process was repeated for a sufficient number of plates to plate out a

total number of plaques at least equal to the library size. These plates were then maintained at 37°C for 5 hours. The plates were then overlaid with nitrocellulose filters that had been pretreated with a solution containing 10 mM isopropyl-beta-D-thiogalactopyanoside (IPTG) and maintained at 37°C for 4 hours. The orientation of the nitrocellulose filters in relation to the plate were marked by punching a hole with a needle dipped in waterproof ink through the filter and into the bacterial plates at several locations. The nitrocellulose filters were removed with forceps and washed once in a TBST solution containing 20 mM Tris-HCl at pH 7.5, 150 mM NaCl and 0.05% monolaurate (tween-20). A second nitrocellulose filter that had also been soaked in a solution containing 10 mM IPTG was reapplied to the bacterial plates to produce duplicate filters. The filters were further washed in a fresh solution of TBST for 15 minutes. Filters were then placed in a blocking solution consisting 20 mM Tris-HCl at pH 7.5, 150 mM NaCl and 1% BSA and agitated for 1 hour at room temperature. The nitrocellulose filters were transferred to a fresh blocking solution containing a 1 to 500 dilution of the primary antibody and gently agitated for at least 1 hour at room temperature. After the filters were agitated in the solution containing the primary antibody the filters were washed 3 to 5 times in TBST for 5 minutes each time to remove any of the residual unbound primary antibody. The filters were transferred into a solution containing fresh blocking solution and a 1 to 500 to a 1 to 1,000 dilution of alkaline phosphatase conjugated secondary antibody. The filters were gently agitated in the solution for at least 1 hour at room temperature. The filters were washed 3 to 5 times in a solution of TBST for at least 5 minutes each time to remove any residual unbound secondary antibody. The filters were washed once in a solution containing 20 mM Tris-HCl at pH 7.5 and 150 mM NaCl. The filters were removed from this solution and excess moisture blotted

from them with filter paper. The color was developed by placing the filter in a solution containing 100 mM Tris-HCl at pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.3 mg/ml of nitro Blue Tetrazolium (NBT) and 0.15 mg/ml of 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) for at least 30 minutes at room temperature. The residual color development solution was rinsed from the filter with a solution containing 20 mM Tris-HCl at pH 7.5 and 150 mM NaCl. The filter was then placed in a stop solution consisting of 20 mM Tris-HCl at pH 2.9 and 1 mM EDTA. The development of an intense purple color indicates a positive result. The filters are used to locate the phage plaque that produced the desired protein. That phage plaque is segregated and then grown up for further analysis.

Several different combinations of primary antibodies and second antibodies were used. The first combination used a primary antibody immunospecific for a decapeptide that will be expressed only if the $V_H$ antigen binding protein is expressed in the proper reading frame to allow read through translation to include the decapeptide epitope covalently attached to the $V_H$ antigen binding protein. This decapeptide epitope and an antibody immunospecific for this decapeptide epitope was described by Green et al., Cell 28:477 (1982) and Niman et al., Proc. Nat. Acad. Sci. U.S.A. 80:4949 (1983). The sequence of the decapeptide recognized is shown in Figure 11. A functional equivalent of the monoclonal antibody that is immunospecific for the decapeptide can be prepared according to the methods of Green et al. and Niman et al. The secondary antibody used with this primary antibody was a goat antimouse IgG (Fisher Scientific). This antibody in immunospecific for the strand region of mouse IgG and did not recognize any portion of the variable region of heavy chain. This particular combination of primary and secondary antibodies when used according to the above protocol determined that between 25% and 30% of the clones were

expressing the decapeptide and therefore these clones were assumed to also be expressing a $V_H$ antigen binding protein.

In another combination the anti-decapeptide mouse monoclonal was used as the primary antibody and an affinity purified goat anti-mouse Ig, commercially available as part of the picoBlue immunoscreening kit from Stratagene Cloning System, La Jolla, CA, was use as the secondary antibody. This combination resulted in a large number of false positive clones because the secondary antibody also immunoreacted with the $V_H$ of the heavy chain. Therefore this antibody reacted with all clones expressing any $V_H$ protein and this combination of primary and secondary antibodies did not specifically detect clones with the $V_H$ polynucleotide in the proper reading frame and thus allowing expressing of the decapeptide.

Several combinations of primary and secondary antibodies are used where the primary antibody is conjugated to fluorescein isothiocyanate (FITC) and thus the immunospecificity of the antibody was not important because the antibody is conjugated to the preselected antigen (FITC) and it is that antigen that should be bound by the $V_H$ antigen binding proteins produced by the clones in the expression library. After this primary antibody has bound by virtue that is FITC conjugated mouse monoclonal antibody p2 5764 (ATCC #HB-9505). The secondary antibody used with this primary antibody is a goat anti-mouse $Ig^6$ (Fisher Scientific, Pittsburgh, PA) conjugated to alkaline phosphatase using the method described in Antibodies: A Laboratory Manual, Harlow and Lowe, eds., Cold Spring Harbor, New York, (1988). If a particular clone in the $V_H$ expression, library, expresses a $V_H$ binding protein that binds the FITC covalently coupled to the primary antibody, the secondary antibody binds specifically and when developed the alkaline phosphate causes a distinct purple color to form.

The second combination of antibodies of the type uses a primary antibody that is FITC conjugated rabbit anti-

human IgG (Fisher Scientific, Pittsburgh, PA). The secondary antibody used with this primary antibody is a goat anti-rabbit IgG conjugated to alkaline phosphatase using the methods described in Antibodies A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, New York, (1988). If a particular clone in the $V_H$ expression library expresses a $V_H$ binding protein that binds the FITC conjugated to the primary antibody, the secondary antibody binds specifically and when developed the alkaline phosphatase causes a distinct purple color to form.

Another primary antibody was the mouse monoclonal antibody p2 5764 (ATCC # HB-9505) conjugated to both FITC and $^{125}I$. The antibody would be bound by any $V_H$ antigen binding proteins expressed. Then because the antibody is also labeled with $^{125}I$, an autoradiogram of the filter is made instead of using a secondary antibody that is conjugated to alkaline phosphatase. This direct production of an autoradiogram allows segregation of the clones in the library expressing a $V_H$ antigen binding protein of interest.

17. Segregating DNA Homologs For a $V_H$ and $V_L$ that Form an Antigen Binding $F_v$

To segregate the individual clones containing DNA homologs that code for a $V_H$ and $V_L$ that form an antigen binding $F_v$, an $V_H$ and $V_L$ expression library was titred according to Example 15. The titred expression library was then screened for the presence of the decapeptide tag expressed with the $V_H$ using the methods described in Example 16. DNA was then prepared from the clones to express the decapepide tag. This DNA was digested with the restriction endonuclease Pvu II to determine whether these clones also contained a $V_L$ DNA homolog. The slower migration of a PvuII restriction endonuclease fragment indicated that the particular clone contained both a $V_H$ and a $V_L$ DNA homolog.

The clones containing both a $V_H$ and a $V_L$ DNA homolog were analyzed to determine whether these clones produced an assembled $F_v$ protein molecule from the $V_H$ and $V_L$ DNA homologs.

The $F_v$ protein fragment produced in clones containing both $V_H$ and $V_L$ was visualized by immune precipitation of radiolabeled protein expressed in the clones. A 50 ml culture of LB broth (5 g/L yeast extract, 10 g/L and tryptone 10 g/L NaCl at pH 7.0) containing 100 ug/ul of ampicillin was inoculated with E. Coli harboring a plasmid contain a $V_H$ and a $V_L$. The culture was maintained at 37°C with shaking until the optical density measured at 550 nm was 0.5. The culture then was centrifuged at 3,000 g for 10 minutes and resuspended in 50 ml of M9 media (6 g/L $Na_2HPO_4$, 3 g/L $KH_2PO_4$, 0.5 g/L NaCl, 1 g/L $NH_4Cl$, 2g/L glucose, 2 mM $MgSO_4$ and 0.1 mM$MgSO_4$ $CaCl_2$ supplemented with amino acids without methionine or cysteine. This solution was maintained at 37°C for 5 minutes and then 0.5 mCi of $^{35}S$ as $HSO_4$ (New England Nuclear, Boston, MA) was added and the solution was further maintained at #&C for an additional 2 hours. The solution was then centrifuged at 300 x g and the supernatant discarded. The resulting bacterial cell pellet was frozen and thawed and then resuspended for 10 minutes and the resulting pellet discarded. The supernatant was admixed with 10 ul of anti-decapeptide monoclonal antibody and maintained for 30-90 minutes on ice. 40 ul of protein G coupled to sepharose beads (Pharmacia, Piscataway, NJ) was admixed to the solution and the added solution maintained for 30 minutes on ice to allow an immune precipitate to form. The solution was centrifuged at 10,000 x g for 10 minutes and the resulting pellet was resuspended in 1 ml of a solution containing 100 mM Tris-HCl at Ph 7.5 and centrifuged at 10,000 x g for 10 minutes. This procedure was repeated twice. The resulting immune precipitate pellet was loaded onto a PhastGel Homogenous 20 gel (Pharmacia, Piscataway, NJ)

according to the manufacturer's directions. The gel was dried and used to expose X-ray film.

The resulting autoradiogram is shown in Figure 12. The presence of $V_L$ that was immunoprecipitated because it was attached to the $V_H$-decapepide tag recognized by the precipitating antibody.

18. Generation of a Combinatorial Library of the Immunoglobulin Repertoire in Phage

Vectors suitable for expression of $V_H$, $V_L$, Fv and Fab sequences are diagrammed in Figures 7 and 9. As previously discussed, the vectors were constructed by modification of Lambda Zap by inserting synthetic oligonucleotides into the multiple cloning site. The vectors were designed to be antisymmetric with respect to the Not I and EcoR I restriction sites which flank the cloning and expression sequences. As described below, this antisymmetry in the placement of restriction sites in a linear vector like bacteriophage allows a library expressing light chains to be combined with one expressing heavy chains to construct combinatorial Fab expression libraries. Lambda Zap II $V_L$II (Figure 9) is designed to serve as a cloning vector for light chain fragments and Lambda Zap II $V_H$ (Figure 7) is designed to serve as a cloning vector for heavy chain sequences in the initial step of library construction. These vectors are engineered to efficiently clone the products of PCR amplification with specific restriction sites incorporated at each end.

A. PCR Amplification of Antibody Fragments

The PCR amplification of mRNA isolated from spleen cells with oligonucleotides which incorporate restriction sites into the ends of the amplified product can be used to clone and express heavy chain sequences including Fd and kappa chain sequences. The oligonucleotides primers used for these amplifications are presented in Tables 1

and 2. The primers are analogous to those which have been successfully used in Example 6 for amplifications of $V_H$ sequences. The set of 5' primers for heavy chain amplification were identical to those previously used to amplify $V_H$ and those for light chain amplification were chosen on similar principles, Sastry et al., Proc. Natl. Acad. Sci. USA, 8G: 5728 (1989) and Orland et al., Proc Natl. Acad. Sci. USA, 8G:3833 (1989). The unique 3' primers of heavy (IgG1) and light (k) chain sequences were chosen to include the cysteines involved in heavy-light chain disulfide bond formation. At this stage no primer was constructed to amplify lambda light chains since they constitute only a small fraction of murine antibodies. In addition, Fv fragments have been constructed using a 3' primer which is complementary to the to the mRNA in the J (joining) region (amino acid 128) and a set of unique 5' primers which are complementary to the first strand cDNA in the conserved N-terminal region of the processed protein. Restriction endonuclease recognition sequences are incorporated into the primers to allow for the cloning of the amplified fragment into a lambda phage vector in a predetermined reading frame for expression.

B. Library Construction

The construction of a combinatorial library was accomplished in two steps. In the first step, separate heavy and light chain libraries were constructed in Lambda Zap II $V_H$ and Lambda Zap II $V_L$ II respectively. In the second step, these two libraries were combined at the antisymmetric EcoR1 sites present in each vector. This resulted in a library of clones each of which potentially co-expresses a heavy and a light chain. The actual combinations are random and do not necessarily reflect the combinations present in the B-cell population in the parent animal. Lambda Zap II $V_H$ expression vector has been used to create a library of heavy chain sequences for DNA obtained by PCR amplifications of mRNA isolated from the

spleen of a 129 $G_{ix}$ + mouse previously immunized with p-nitrophenyl phosphonamidate (NPN) antigen 1 according to formula I (Figure 13) conjugated to keyhole limpet hemocyanin (KLH).

The NPN-KLH conjugate was prepared by admixture of 250 ul of a solution containing 2.5 mg of NPN according to formula 1 (Figure 12) in dimethylformamide with 750 ul of a solution containing 2 mg of KLH in 0.01 M sodium phosphate buffer (pH 7.2). The two solutions were admixed by slow addition of the NPN solution to the KLH solution while the KLH solution was being agitated by a rotating stirring bar. Thereafter the admixture was maintained at 4°C for 1 hour with the same agitation to allow conjugation to proceed. The conjugated NPN-KHL was isolated from the nonconjugated NPN and KLH by gel filtration through Sephadex G-25. The isolated NPN-KLH conjugate was used in mouse immunizations as described in Example 3.

The spleen mRNA resulting from the above immunizations was isolated and used to create a primary library of $V_H$ gene sequences using the Lambda Zap II $V_H$ expression vector. The primary library contains 1.3 x $10^6$ pfu and has been screened for the expression of the decapeptide tag to determine the percentage of clones expressing Fd sequences. The sequence for this peptide is only in frame for expression following the cloning of an Fd (or $V_H$) fragment into the vector. At least 80% of the clones in the library express Fd fragments based on immuno-detection of the decapeptide tag.

The light chain library was constructed in the same way as the heavy chain and shown to contain 2.5 x $10^6$ members. Plaque screening, using the anti-kappa chain antibody, indicated that 60% of the library contained expressed light chain inserts. This relatively small percentage of inserts probably resulted from incomplete dephosphorylation of vector after cleavage with Sac I and Xba I.

Once obtained, the two libraries were used to construct a combinatorial library by crossing them at the EcoRI site. To accomplish the cross, DNA was first purified from each library. The light chain library was cleaved with MluI restriction endonuclease, the resulting 5' ends dephosphorylated and the product digested with EcoRI. This process cleaved the left arm of the vector into several pieces but the light arm containing the light chain sequences, remained intact. In a parallel fashion, the DNA of heavy chain library was cleaved with HindIII, dephosphorylated and cleaved with EcoR I, destroying the right arm but leaving the left arm containing the heavy chain sequences intact. The DNA's so prepared were then combined and ligated. After ligation only clones which resulted from combination of a right arm of light chain-containing clones reconstituted a viable phage. After ligation and packaging, $2.5 \times 10^7$ clones were obtained. This is the combinatorial Fab expression library that was screened to identify clones having affinity for NPN. To determine the frequency the phage clones which co-express the light and heavy chain fragments, duplicate lifts o the light chain, heavy chain and combinatorial libraries were screened as above for light and heavy chain expression. In this study of approximately 500 recombinant phage approximately 60% co-expressed light and heavy chain proteins.

C. Antigen Binding

All three libraries, the light chain, the heavy chain and Fab were screened to determine if they contained recombinant phage that expressed antibody fragments binding NPN. In a typical procedure 30,000 phage were plated and duplicate lifts with nitrocellulose screened for binding to NPN coupled to $^{125}I$ labeled BSA (Figure 15). Duplicate screens of 80,000 recombinant phage from the light chain library and a similar number from the heavy chain library did not identify any clones which bound the

antigen. In contrast, the screen of a similar number of clones from the Fab expression library identified many phage plaques that bound NPN (Figure 15). This observation indicates that under conditions where many heavy chains in combination with light chains bind to antigen the same heavy or light chains alone do not. Therefore, in the case of NPN, it is believed that there are many heavy and light chains that only bind antigen when they are combined with specific light and heavy chains respectively.

To assess the ability to screen large numbers of clones and obtain a more quantitative estimate of the frequency of antigen binding clones in the combinatorial library, one million phage plaques were screened and approximately 100 clones which bound to antigen were identified. For six clones which were believed to bind NPN, a region of the plate containing the positive and approximately 20 surrounding bacteriophage plaques was "cored", replated, and screened with duplicate lifts (Figures 15). As expected, approximately one in twenty of the phage specifically bind to antigen. "Cores" of regions of the plated phage believed to be negative did not give positives on replating.

To determine the specificity of the antigen-antibody interaction, antigen binding was competed with free unlabeled antigen as shown in Figure 16. Competition studies showed that individual clones could be distinguished on the basis of antigen affinity. The concentration of free hapten required for complete inhibition of binding varied between 10-100 x $10^{-9}$ M suggestion that the expressed Fab fragments had binding constants in the nanomolar range.

D. Composition of the Clones and Their Expressed Products

In preparation for characterization of the protein products able to bind NPN as described in Example 19C, a

plasmid containing the heavy and light chain genes was excised from the appropriate "cored" bacteriophage plaque using M13mp8 helper phage. Mapping of the excised plasmid demonstrated a restriction pattern consistent with incorporation of heavy and light chain sequences. The protein products of one of the clones was analyzed by ELISA and Western blotting to establish the composition of the NPN binding protein. A bacterial supernate following IPTG induction was concentrated and subjected to gel filtration. Fractions in the molecular weight range 40-60 kD were pooled, concentrated and subjected to a further gel filtration separation. As illustrated in Figure 17, ELISA analysis of the eluting fractions demonstrated that NPN binding was associated with a protein of molecular weight about 50 kD which immunological detection showed contained both heavy and light chains. A Western blot (not shown) of a concentrated bacterial supernate preparation under non-reducing conditions was developed with anti-decapeptide antibody. This revealed a protein band of molecular weight of 50 kD. Taken together these results are consistent with NPN binding being a function of Fab fragments in which heavy and light chains are covalently linked.

20. Flp recombinase-catalyzed Recombination

Experiments directed to the in vivo recombination of two lambda vectors using flp recombinase-catalyzed recombination are described. The flp recombination site was introduced into the phage vectors using 39mer synthetic oligonucleotides. The sequence of the flp site utilized for recombination was derived from several references (e.g. Senecoff et al., Proc. Nat. Acad. Sci. USA 82:7220-7224 (1985)). The XbaI site within the 8bp core was eliminated as this site was to be used in the cloning strategy. This was accomplished by making a point mutation which has little or no effect on its ability to allow recombination (McLeod et al., Mol. Cell. Biol. 6:3357-

3367 (1985)). However, this point mutation is not required for the system to function. The oligonucleotides were further designed to be inserted in the EcoR1 sites of the Lambda Zap II $V_H$ and Lambda Zap II $V_L$ vectors so that only one flanking EcoR1 sites would be regenerated (see Figure 18). The flanking sequences are not essential to the system.

The following sequences were inserted into Lamba Zap II $V_H$:

(110) Oligo 79 EcoR1
AATTCGAAGTTCCTATTCTCTAAAAAGTATAGGAACTTC 3'

(111) Oligo 80 GCTTCAAGGATAAGAGATTTTTCATATCCTTGAAGTTAA 5'

The following sequences were inserted into Lambda Zap II $V_L$:

(112) Oligo 81
AATTGAAGTTCCTATTCTCTAAAAAGTATAGGAACTTCG EcoR1 3'

(113) Oligo 82 CTTCAAGGATAAGAGATTTTTCATATCCTTGAAGCTTAA 5'

Vectors were constructed as follows. The first two oligonucleotides were mixed (0.5 $\mu$g oligo 79, 0.5 $\mu$g oligo 80, 1 $\mu$l 200 mM Tris, pH 7.4, 20 mM MgCl, 500 mM NaCl, and $H_2O$ to 10 $\mu$l), heated to 85°C 5 min., and allowed to cool to room temperature over 1 hour in a water bath. The procedure was repeated using oligos 81 and 82.

Ligation into vector arms was accomplished by digesting Lambda Zap $V_H$ and Lambda Zap II $V_L$ with 3U/$\mu$g EcoR1 according to standard digestion procedure. After phenol/chloroform extraction, DNA was precipitated with EtOH. The vector was not phosphatase treated so that the oligonucleotides could be inserted without kinase treatment, thus preventing multiple tandem oligonucleotide inserts. Ligations were performed in 5 $\mu$l volumes using 1 $\mu$g of lambda DNA and 1 ng of annealed oligonucleotides according to standard ligation protocol (see Maniatis et al., supra). Ligation mixes were packaged using Gigapack

Gold™ (Stratagene Cloning Systems, San Diego, CA) according to the protocol recommended in the manual.

Following packaging, the vectors were screened. Packaged DNA was plated according to the Gigapack Gold™ manual procedure on NZY agar with approximately 400 pfu per 100 mm Petri dish. Duplicate plaque lifts were done according to the protocol in the Predigested ZapII Cloning Manual (Stratagene Cloning Systems, San Diego, CA) on nitrocellulose filters. Denaturation and fixation of DNA onto the membranes is also described in the manual. Prehybridization was performed according to pBluescript II Exo/Mung DNA Sequencing System™ instruction manual (Stratagene Cloning Systems, San Diego, CA) for oligonucleotide probes (pg 6). Hybridization was performed overnight using $^{32}P$ kinased oligo 79 ($0.5 \times 10^6$ cpm/ml) according to the pBluescript manual (Stratagene Cloning Systems, San Diego, CA). Oligo 79 was kinased using standard $^{32}P$ gamma ATP labelling techniques (see Maniatis et al., supra). Filters were washed in 6X SSC, 0.1% SDS, 3 times at room temperature, once at 55°C and finally at 59°C. Each was washed for approximately 10 minutes. Positive plaques were identified using X-ray autoradiography. Twelve duplicate plaques were cored in 500 µl SM, 20 µl chloroform. These plaques were sufficiently well isolated that secondary screening was not required. The cored plaques were excised according to the Predigested Zap II Cloning Manual (Stratagene Cloning Systems, San Diego, CA) and DNA from single ampicillin resistant colonies was sequenced using minipreped DNA and the T7 and T3 primers according to the DSK 35S Sequencing kit (Stratagene Cloning Systems, San Diego, CA). Clones with flp sites in the correct orientation and opposite orientation were identified, amplified and titred. One of each type of clone (FLPHC+, FLPHC-, FLPLC+, FLPLC-) was used to test in vivo flp- mediated recombination.

In vivo flp-mediated recombination was accomplished as follows. Flp recombinase was expressed off the tac

promoter on a plasmid, pCS3, in E. coli MM294 strain (Lebreton et al. 1988). This is a low copy number plasmid with the pACYC origin of replication and contains a chloramphenicol resistance gene.

5 x $10^8$ cells were coinfected with FLPHC and FLPLC vectors at an moi of 5 and 10 pfu each per cell. Combinations of FLPHC+ and FLPLV+, or FLPHC- and FLPLC-, or FLPHC+ and FLPLC- were tested.

Overnight cultures of MM294(pCS3) were grown in LB, spun down and resuspended in 10mM $MgSO_4$ at a density of $OD_{600}$ = 1.0. The appropriate amounts of phage were added to 0.5 ml of cells and allowed to adhere at 37°C for 15 minutes. 50 ml of NZY was added to each flask and incubated for 2 hours with shaking. 250 $\mu$l of chloroform was added to 25 ml of lysate and incubated for 15 minutes at room temperature. The supernatants were titred and screened for phage containing both Lambda Zap II $V_H$ left arms and Lambda Zap II $V_L$ right arms. Probes to identify Zap II $V_H$ left arms and Lambda Zap II $V_L$ right arms were designed by identifying unique sequences from the known sequence of the vectors.

The Lambda Zap II $V_H$ left arm probe had the following sequence:

(114) CTAGTTACCCGTACGACCCCCCCGTTCCGGACTACGCTTCTTAATAG 3'

This sequence hybridizes to the decapeptide sequence of the Lambda Zap II $V_H$. The Lambda Zap II $V_L$ right arm probe had the following sequence:

(115) 5' GAGCTCGTCAGTTCTAGAGTTAAGCGGCCG 3'

This sequence hybridizes to the sequence from the Sac1 site to the former Not1 site of the Lambda Zap II $V_L$ vector.

The screening procedure used was the same as that used to identify the flp vectors, as described above, with the exception of washing conditions. Filters were washed with 6XSSC, 1%SDS 3 times at room temperature and twice at 60°C. Plaques which hybridized to both probes were identified by X-ray autoradiography, cored, excised and

digested to determine if recombination had occurred. Control plaques identified as hybridizing to only one probe and to neither probe were also cored. Diagnostic restriction digests were PvuII, Pvul, Xhol, Xhol/Pvul, Sac1, Sac/PVul, Not1, Xbal, ScaI, SpeI, SpeI/PuuI. Restriction digest results verified that recombination at the flp site occurred in vivo in cells expressing the flp recombinase gene and not in control SURE™ E. coli cells (Stratagene Cloning Systems, San Diego, CA) which do not normally express flp recombinase.

Efficiency of recombination according to the number of plaques identified as hybridizing to both probes was initially between about 5-10%. Changes to the protocol can be made, however, which will improve the efficiency of recovery of recombined vectors. For example, by adding selectable marker sequences to the left and right arms of the vectors, up to 100% of target recombinants can be identified (Figure 20). Adding selection systems to ensure that all recombinants contain inserts will also increase the efficiency of identifying the desired clones.

In Example 19 a relatively restricted library was prepared because only a limited number of primers were used for PCR amplification of Fd sequences. The library is expected to contain only clones expressing kappa/gamma sequences. However, this is not an inherent limitation of the method since additional primers can be added to amplify any antibody class or subclass. Despite this restriction we were able to isolate a large number of antigen binding clones. Of interest is how a phage library prepared as described herein compares with the in vivo antibody repertoire in terms of size, characteristics of diversity, and ease of access.

The size of the mammalian antibody repertoire is difficult to judge but a figure of the order of $10^6$-$10^8$ different antigen specificities is often quoted. With some of the reservations discussed below, a phage library of this size or large can readily be constructed by a

modification of the current method. In fact once an initial combinatorial library has been constructed, heavy and light chains can be shuffled to obtain libraries of exceptionally large numbers.

In principle, the diversity characteristics of the naive (unimmunized) in vivo repertoire and corresponding phage library are expected to be similar in that both involve a random combination of heavy and light chains. However, different factors will act to restrict the diversity expressed by an in vivo repertoire and phage library. For example a physiological modification such as tolerance will restrict the expression of certain antigenic specificities from the in vivo repertoire but these specificities may still appear in the phage library. For example, the representation of mRNA for sequences expressed by stimulated B-cells can be expected to predominate over those of unstimulated cells because of higher levels of expression. Different source tissues (e.g., peripheral blood, bone marrow or regional lymph nodes) and different PCR primers (e.g., ones expected to amplify different antibody classes) may result in library with different diversity characteristics.

Another difference between in vivo repertoire and phage library is that antibodies isolated from the former may have benefited from affinity maturation due to somatic mutations after combination of heavy and light chains whereas the latter randomly combines the matured heavy and light chains. Given a large enough phage library derived from a particular in vivo repertoire, the original matured heavy and light chains will be recombined. However, since one of the potential benefits of this new technology is to obviate the need for immunization by the generation of a single highly diverse "generic" phage library, it would be useful to have methods to optimize sequences to compensate for the absence of somatic mutation and clonal selection. Three procedures are made readily available through the methods of the present invention. First, saturation muta-

genesis may be performed on the CDR's and the resulting Fabs can be assayed for increased function. Second, a heavy or a light chain of a clone which binds antigen can be recombined with the entire light or heavy chain libraries respectively in a procedure identical to the one used to construct the combinatorial library. Third, iterative cycles of the two above procedures can be performed to further optimize the affinity or catalytic properties of the immunoglobulin. It should be noted that the latter two procedures are not permitted in B-cell clonal selection which suggests that the methods described here may actually increase the ability to identify optimal sequences.

Access is the third area where it is of interest to compare the in vivo antibody repertoire and phage library. In practical terms the phage library is much easier to access. The screening methods allow one to survey at least 50,000 clones per plate so that $10^6$ antibodies can be readily examined in a day. This factor alone should encourage the replacement of hybridoma technology with the methods described here. The most powerful screening methods utilize selection which may be accomplished by incorporating selectable markers into the antigen such as leaving groups necessary for replication of auxotrophic bacterial strains or toxic substituents susceptible to catalytic inactivation. There are also further advantages related to the fact that the in vivo antibody repertoire can only be accessed via immunization which is a selection on the basis of binding affinity. The phage library is not similarly restricted. For example, the only general method to identify antibodies with catalytic properties has been by pre-selection on the basis of affinity of the antibody to a transition state analogue. No such restrictions apply to the in vivo library where catalysis can, in principle, be assayed directly. The ability to directly assay large numbers of antibodies for function may allow selection for catalysts in reactions where a mechanism is

not well defined or synthesis of the transition state analog is difficult. Assaying for catalysis directly eliminates the bias of the screening procedure for reaction mechanisms pejorative to a synthetic analog and therefore simultaneous exploration of multiple reaction pathways for a given chemical transformation are possible.

Although we have given examples of several screening methods, it should be clear to one skilled in the art that alternative methods of screening, usch as by panning clels or particles expressing the protein product on their surface would essentially be equivalent. If the expressed gene products of interest are RNA molecules instead of proteins, screening could be accomplished by nucleic acid hybridization or by detecting some functional property of th eRNA, such as ribozyme catalysis.

The methods disclosed herein describe generation of Fab fragments which are clearly different in a number of important respects from intact (whole) antibodies. There is undoubtedly a loss of a affinity in having monovalent Fab antigen binders but this can be compensated by selection of suitably tight binders. For a number of applications such as diagnostics and biosensors it may be preferable to have monovalent Fab fragments. For applications requiring Fc effector functions, the technology already exists for extending the heavy chain gene and expressing the glycosylated whole antibody in mammalian cells.

The ideas presented here address the bottle neck in the identification and evaluation of antibodies. It is now possible to construct and screen at least three orders of magnitude more clones with mono-specificity than previously possible. The potential applications of the method should span basic research and applied sciences.

21. Oligonucleotide Primer Design for Producing Dicistronic DNA

A method based on PCR amplification that fuses heavy and light chain sequences has been used to construct a

complete antigen binding domain of a Fab protein fragment composed of a heavy and a light chain. Schematic diagrams of an immunoglobulin molecule composed of heavy and light chains containing constant and variable regions is shown in Figure 1. Human heavy chain IgG and human kappa light chain are diagrammatically sketched in Figures 2A and 2B, respectively. To accomplish this procedure, immunoglobulin heavy and light chain primers were designed to produce a region of homology between two polymerase chain reaction (PCR) products. The complementary regions have been shown to hybridize predominantly under conditions where one set of primers ("inside primer pair") is used in a limiting amount relative to the other set of primers ("outside primer pair"). After the 3' ends of the PCR products have hybridized, the DNA polymerase has been shown to extend the ends creating a fusion sequence carrying the unique sequences of both PCR fragments separated by one copy of region X cistronic bridge. A two-step cloning procedure is thus avoided. When the recombinant sequence is then inserted into an expression vector such as ImmunoZAP, a fusion production capable of simultaneously expressing the heavy and light chains can be produced.

The strategy used for producing immunoglobulin heavy and light chain PCR dicistronic DNA is shown schematically in Figure 21. Regions of the immunoglobulin heavy chain coding strand are designated $V_H$, $C_H1$, $C_H2$, and $C_H3$ corresponding to functional regions in the protein. The corresponding regions of the non-coding strand are designated by a prime ('). Regions $V_L$ and $C_L$ are similarly labelled for the kappa light chain. This procedure can also be performed using lambda light chain specific regions. A region, X, unrelated to the natural immunoglobulin sequences, is introduced into the fusion product by attaching X to the 5' ends of both of the $C_H1'$ and $V_L$ inside primers.

Overlapping oligonucleotide primers used in the fusion-PCR reactions to produce dicistronic DNA were

designed to encode the following: amino acids of 225 to 230 of the IgG heavy chain hinge region which are common to all human IgG isotypes; an Spe I restriction site; two stop codons; a ribosome binding site; a periplasmic (pelB) leader sequence (Better, et al., Science, 240:1041-1043 (1988); Lei, et al., J. Bacteriol., 169:4379-4383 (1988)); a Sac I restriction site which encodes amino acids 1 and 2 of the mature kappa light chain; and amino acids 3 to 8 of the mature kappa light chain. The X region was designed to contain a ribosome binding site and a pelB leader to ensure expression of the light chain. Nucleotide sequences for all human and mouse PCR primers, both inside and outside, are listed in Table 11. Primers followed by a prime (') represent non-coding strand sequences.

Table 11

Human and Mouse PCR Primers

| Seq. Id. No. | | | |
|---|---|---|---|
| | Human | | |
| (117) | $V_H$ | 5'-GTCCTGTCCGAGGTGCAGCTGCTCGAGTCTGG-3' | |
| (118) | $C_H1'$ | 5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGGT<br>ATTTCATTATGACTGTCTCCTTGCTATTAACTAG<br>TACAAGATTTGGGCTC-3' | |
| (119) | $V_L$ | 5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGCT<br>GCCCAACCTGCCATGGCTGAGCTCGTGATGACCC<br>CAGTCTCC-3' | |
| (120) | $C_L'$ | 5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA<br>GCTCTTTGTGACGGGCGAACTC-3' | |
| | Mouse | | |
| (121) | $V_H$ | 5'-AGGTCCAGCTGCTCGAGTCTGG-3' | |
| (122) | $C_H1'$ | 5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGG<br>TATTTCATTATGACTGTCTCCTTGCTATTAACT<br>AGTATACAATCCCTGGGCACAAT-3' | |
| (123) | $V_L$ | 5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGC<br>TGCCCAACCTGCCATGGCTGAGCTCGTGATGAC<br>CCAGTCTCC-3' | |

(124) $C_L'$ 5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA-3'

---

The overlapping regions of the human $C_H1'$ inside and $V_L$ inside primers are illustrated in Figure 22. The heavy chain downstream $C_H1'$ inside primer sequence is written 3' to 5' and the light chain upstream $V_L$ inside primer sequence is written 5' to 3'. The complementary PCR product strands, and not the primer strands, cross-prime to create the dicistronic molecule. Bold nucleotides represent regions where the $C_H1'$ inside primer hybridizes to the 3' end of $C_H1$ on human IgG heavy chain mRNA or where the $V_L$ inside primer hybridizes to the 5' end of $V_L$ framework on human kappa light chain cDNA. The amino acid and nucleotides in italics represent changes in sequence from the original pelB leader sequence.

At amino acid 15 of the pelB leader sequence, the codon was changed from CTC to ATC resulting in a conservative amino acid change from a leucine to an isoleucine as shown in Figure 22 and Table 11. Hydrophobic amino acids in the core region of periplasmic leader sequences have been shown to be essential for correct processing of the leader sequence and transport of the mature protein to the periplasm. Oliver, in Neidhardt, R.C. (ed.), Escherichia coli and Salmonella Typhimurium., Am. Soc. Microbiol., 1:56-69 (1987). The nucleotide changes were made to allow for the artifactual insertion of one or two dATPs at the 3' end of the overlapping dicistronic molecules. Thermus aquaticus (Taq) DNA polymerase may add a dATP to the 3' end of the PCR product because of terminal transferase activity. Jiang, etg al. Oncogene, 4:923-928 (1989). The additional dATP would then cause a mismatch between the overlapping PCR products at the 3' terminus and inhibit elongation by Taq DNA polymerase. Sommer, et al. Nucl. Acids Res., 17:6749 (1989). Therefore, the change to two dTTPs in this position of the oligonucleotide primers would allow proper base pairing if up to two dATPs were added to the 3' terminus of the heavy chain PCR product.

The kappa light chain PCR product was designed to terminate at a position where two dTTPs occur 5' of the end of the product and did not require alterations of the nucleotide sequence. Nucleotides were changed in the kappa light chain primer encoding the pelB leader sequence without introducing amino acid changes in order to decrease the number of mismatches between the primer and the leader sequence of the kappa light chain mRNA as shown in Figure 22 and Table 11.

All primers were synthesized on an Applied Biosystems DNA synthesizer, Model 381A, following the manufacturer's instructions.

22. Preparation of a $V_H$-and $V_L$-Coding Repertoire

A. Preparation of a $V_H$-and $V_L$-Coding REpertoire from a Human cDNA Combinatorial Library

Cloned DNA, previously isolated from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT) was used as a template for preparing a $V_H$-and $V_L$-coding repertoire. Mullinax, et al., supra. Briefly, the combinatorial library was prepared by the following approach. Volunteer donors, who had been previously immunized against tetanus but had not received booster injections within the last year, received injections on 2 consecutive days of 0.5 milliliters (ml) of alum-absorbed tetnus toxoid (TT) (40 microgram/ml (ug)/ml) (Connaught Laboratories, Swiftwater, Pennsylvania).

One hundred ml of blood was drawn from the volunteers 6 days post injection and anticoagulated with a mixture of 0.14 M citric acid, 0.2 M trisodium citrate, and 0.22 M dextrose. The peripheral blood lymphocytes (PBLs) were recovered and isolated from the whole blood by layering the whole blood on Histopaque-1077 (Sigma, St. Louis, Missouri) and centrifuging at 400 x g for 30 minutes at 25 degrees Celsius (25°C). Isolated PBLs were washed twice with phosphate buffered saline (PBS) (150 mM sodium chloride and 150 mM sodium phosphate, pH 7.2 at 25°C).

Total RNA was then purified from the PBLs ($10^6$ B cells per ml blood per 100 ml of blood) for an enriched source of B-cell mRNA coding for antiTT IgG using an RNA isolation kit according to manufacturer's instructions (Stratagene, La Jolla, California) and also described by Chomczynski et al., Anal. Biochem., 162:156-159 (1987). Briefly, the isolated PBLs were homogenized in 10 ml of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M beta-mercaptoethanol. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized cells. Ten ml of phenol that had been previously saturated with $H_2O$ was also admixed to the denaturing solution containing the homogenized cells. Two ml of a chloroform: isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuged tube (Fisher Scientific Company, Pittsburgh, Pennsylvania). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described above. Three ml of isopropyl alcohol was added to the re-suspended total cellular RNA and inverted to mix. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for

15 minutes and then re-suspended in diethyl pyrocarbonate (DEPC) treated (DEPC-$H_2O$) $H_2O$).

Messenger RNA (mRNA) was prepared from the total cellular RNA using methods described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY, (1982). Briefly, 500 mg of the total RNA isolated from a PBLs prepared as described above was re-suspended in one ml of 1X sample buffer (1 mM Tris-HCl, (Tris [hydroxylmethyl-aminomethane]) pH 7.5; 0.1 mM EDTA (disodium ethylene diamine tetra-acetic acid), 0.5 M NaCl) and maintained at 65°C for five minutes and then on ice for five more minutes. The mixture was then applied to an oligo-dT (Stratagene) column that was previously prepared by washing the oligo-dT with a solution containing 10 mM Tris-HCl, pH 7.5; 1 mM EDTA, 0.5 M NaCl. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for five minutes at 65°C. The oligo dT column was then washed with 0.4 ml of high salt loading buffer consisting of 10 mM Tris-HCl at pH 7.5, 500 mM sodium chloride, and 1 mM EDTA. The oligo dT column was then washed with 2 ml of 1 X low salt buffer consisting of 10 mM Tris-HCl at pH 7.5, 100 mM sodium chloride, and 1 mM EDTA. The messenger RNA was eluted from the oligo dT column with 0.6 ml of buffer consisting of 10 mM Tris-HCl at pH 7.5, and 1mM EDTA. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and re-suspended in DEPC $H_2O$.

The messenger RNA isolated by the above process contains a plurality of different $V_H$ and $V_L$ coding polynucleotides, i.e., greater than about $10^4$ different $V_H$- and $V_L$-coding genes.

Isolated RNA was converted to cDNA by a primer extension reaction with a first-strand synthesis kit according to manufacturer's instructions (Stratagene) by using an oligo (dT) primer for the light chain and a specific

primer, $C_H1'$, for the heavy chain. Mullinax et al., supra. In a typical 50 $\mu$l transcription reaction, 5 ug of PBL mRNA in water was first hybridized (annealed) with 200 ng (50.0 pmol) of an oligo (dT) primer for the light chain. In a separate reaction, 5 ug of PBL mRNA in water was first hybridized (annealed with 200 ng (20 pmol) of the heavy chain primer, $C_H1'$, at 65°C for five minutes. Subsequently, the mixture was adjusted to 0.5 mM each of dATP, dCTP, dGTP and dTTP, 50 mM Tris-HCl at pH 8.3, 3 mM $MgCl_2$ 75 mM KCl, 10 mM DTT, 20 units of RNase block II (Stratagene), and 20 units of Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), was added and the solution was maintained for 1 hour at 37°C. PCR amplification of the heavy and light chain sequences was done separately using 0.25-0.5 ug of first-strand synthesis product as template with sets of primer pairs using Taq DNA polymerase as described in Example 23.

The PCR amplified light chain DNA fragments were then digested with Sac I and Xba I and ligated into a modified Lambda Zap II vector as prepared in Example 29 to form a light chain ImmunoZap Library (ImmunoZAP L; Stratacyte, La Jolla, California). The PCR amplified heavy chain DNA was digested with Spe I and Sho I and ligated into a different modified Lambda Zap II vector as prepared in Example 27 to form a heavy chain ImmunoZap Library (ImmunoZAP H; Stratacyte). The resulting libraries were amplified and the resulting DNA was packaged into bacteriophage with in vitro packaging extract, Gigapack II gold (Stratagene) and used to infect E. coli strain XL1-Blue (Stratagene).

To construct a library for coexpression, the right art of the heavy chain library phage DNA was digested with Hind III, preserving the left arm of ImmunoZAP H with a heavy chain inserts. The left arm of the light chain library phage DNA was digested with Mlu I resulting in a right arm of ImmunoZAP with kappa light chain inserts. Both products were then digested with EcoRI and ligated to create a combinatorial library that encoded human Fab

fragments including those specific for TT. Mullinax, et al., supra.

Reactive plaques were first identified by binding to tetanus toxoid as described in Example 31. Bacteriophage from purified reactive plaques were then converted to the plasmid format by in vivo excision with R408 helper phage (Stratagene) following methods described in Example 31 and familiar to one skilled in the art. Short, et al., Nucl. Acids. Res., 16:7583-7600 (1988). The resulting purified plasmid DNA encoding heavy and light chain was then used in PCR reactions as described below in Example 23.

B. Preparation of a $V_H$- and $V_L$-Coding Repertoire from mRNA from Tissues and Cells

(i) Human

Purified populations of PBLs, other lymphocytes, and hybridomas which express immunoglobulins including IgG, IgM, IgE, IgD, and IgA are used as sources for isolating mRNA encoding immunoglobulins. PBL's and other immunoglobulin expressing lymphocytes are isolated from either spleen, lymphoid tissue or plasma. Following purification of the cells, total RNA is then purified from these cells using a RNA isolation kit (Stratagene) as described in Example 22a. The purified RNA is then converted to cDNA with a first-strand synthesis kit as described in Example 22a. The resultant cDNA is then used as a template in PCR amplification reactions as described below in Example 23 for the production of dicistronic molecules expressing heavy and light chains.

(ii) Mouse

Populations of cells described above can be isolated from other mammalian sources such as mouse or rabbit. Both mRNA and rearranged DNA can be isolated as described above and used as templates in PCR amplification reactions. cDNA synthesized from mRNA isolated from a mouse anti-human fibronectin hybridoma (ATCC, CRL-1606) was used

as a preferred template for the production of dicistronic molecules expressing heavy and light chain.

c. Preparation of a $V_H$-Coding Repertoire from Rearranged DNA

Rearranged DNA isolated from PBLs, other lymphocytes, and hybridomas which express immunoglobulins can be used to prepare a $V_H$-coding repertoire. The amplification procedure for preparing a $V_H$-coding repertoire using rearranged DNA is performed as described in Example 23.

23. Preparation of DNA Homologs

A. $V_H$-Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 22 from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT), was used as a template for preparing a $V_H$-coding double stranded DNA homolog. Human heavy chain, containing both the $V_H$ and $C_H1$ coding region and designated as Fd, was amplified in a PCR reaction. THe amplification was performed in a 100 ul reaction containing 5 nanograms (ng) of the cloned DNA in PCR buffer consisting of the following: 10 mM Tris-HCl, pH 8.3; 50 mM KCl, 1.5 mM $MgCl_2$; 0.001% (w/v) gelatin; 200 mM of each dNTP; 200 nanomolar (nM) of each primer; and 2.5 units of Taq DNA polymerase. The human $V_H$ outside primer and $C_H1'$ inside primer were used as a PCR primer pair for amplification of the heavy chain (Table 11 and Figure 21). The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle (thermocycle) involved denaturation at 94°C for 1.5 minutes, annealing at 54°C for 2.5 minutes and polynucleotide synthesis by primer extension (elongation) at 72°C for 3.0 minutes followed by a return to the denaturation temperature. The resultant amplified $V_H$-coding DNA homolog containing samples were then gel purified, extracted twice with phenol/chloroform, once with chloro-

form followed by ethanol precipitation and were stored at -70°C in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA.

To verify the amplification of the heavy chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the heavy chain was approximately 730 base pairs as shown in Figure 23. The $V_H$-coding double stranded DNA homologs were then used in subsequence PCR amplification reactions with $V_L$-coding counterparts prepared below for the production of dicistronic DNA molecules having $V_H$ and $V_L$ cistronic portions as illustrated in Example 24.

B. $V_L$-Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 22 from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT), was used as a template for preparing a $V_L$-coding double stranded DNA homolog. Human light chain, containing the entire coding region of kappa light chain ($V_L$ and $C_L$), was amplified using the same PCR conditions described for human heavy chain with the exception that a human $V_L$ inside primer and $C_L$' outside primer were used as the PCR primer pair (Table 11 and Figure 21). The resultant $V_L$-coding double stranded DNA homolog was gel purified and stored as described above.

To verify the amplification of the light chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the light chain was approximately 690 base pairs as shown in Figure 23. The $V_L$-coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with $V_H$-coding counterparts prepared above for the production of dicistronic DNA molecules as illustrated in Example 24.

24. Preparation of Internally-Primed Duplexes of $V_H$- and $V_L$-Coding DNA Homolog

A. Hybridization of $V_H$- with $V_L$-Coding DNA Homologs

The $V_H$- and $V_L$-coding double stranded DNA homologs prepare in Examples 23A and 23B, respectively, were admixed together and denatured at 95°C for 5 minutes to separate the strands of each homolog. The denatured $V_H$- and $V_L$-coding DNA strands in the admixture were then annealed at 54°C for 5 minutes to form a $V_H$- and $V_L$-coding duplex DNA molecule hybridized at the 3' ends at region X of each original homolog. One strand of the X region (cistronic) bridge encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon, and a polypeptide leader (pelB) having a translation initiation codon in the same reading frame as the downstream cistron located downstream from the ribosome binding site.

B. Primer Extension to Produce Dicistronic DNA Molecules

The hybridized recombinant $V_H$- and $V_L$-coding DNA molecule (internally primed duplex) was subjected to primer extension and then amplified with only the $V_H$ and $C_L$' primers following the PCR reaction procedure described in Example 23A. This second PCR reaction is schematically represented in Figure 21. The PCR reaction products were gel electrophoresed to verify the presence of the resultant $V_H$- and $V_L$-coding dicistronic DNA molecules. The expected size of the dicistronic molecule was about 1390 base pairs and is shown in Figure 23. The resultant $V_H$- and $V_L$-coding dicistronic DNA molecules were then ligated into the modified ImmunoZAP H vector (Figures 24A and 24B) for the construction of expression vectors as described in Example 30.

2. Preparation of Mouse Hybridoma $V_H$- and $V_L$-Coding Double Stranded DNA Homologs and Production of Dicistronic DNA Molecules in a Single Amplification Reaction

Mouse hybridoma heavy and light chain cDNA prepared in Example 22B was amplified in a single PCR reaction using the reaction conditions given above with an excess of the outside primers (200 nM concentration of both the mouse $V_H$ primer and $C_L$' primer) and a limiting amount of the inside primers (20 nM concentration of both the mouse $C_H1$' and $V_L$ primer) (Table 11). The resultant mouse heavy and light chain dicistronic molecules were then inserted into a modified ImmunoZAP H for construction of an expression vector as described in Example 30.

26. Preparation of Internally-Primed Duplexes Using a Single Internal Primer that Overlaps Both the $V_H$ and $V_L$ Repertoires

Another approach to producing a library of dicistronic DNA molecules is to use a single internal primer instead of using two separately internal primers. The process of creating a dicistronic molecule comprising an upstream $V_H$ cistron and a downstream $V_L$ cistron is to combine in a PCR buffer the following: a repertoire of $V_H$ genes consisting of at least $10^5$ different genes; a repertoire of $V_L$ genes consisting of at least $10^4$ different genes; an outside $V_H$ primer; an outside $V_L$; and a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5' terminal primer-template portion. The PCR reaction is performed as described in Example 22A.

The 3'-terminal priming portion of a polynucleotide strand (linker) has a nucleotide base sequence homologous to a portion of the primer extension product of one of the outside primers. The 5'-terminal priming portion encodes a nucleotide base sequence homologous to a portion of the primer extension product of the other outside primer. The

cistronic bridge coding portion encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon and a polypeptide leader (pelB) having a translation initiation codon in the same reading frame as the downstream cistron where the initiation codon is located downstream from the ribosome binding site. Polynucleotide strand (linker) primers useful in this invention are listed in Table 12.

Table 12

Polynucleotide Strand (Linker) Primers

| Seq. Id. No. | | |
|---|---|---|
| (1251)[1] | 1' | 5' GGAGAGTGGGTCATCACGAGCTCAGCCATGGCAGGTTGG GCAGCGATTAATAACAATCCAGCGGCTGCCGTAGGCAAT AGGTATTTCATTATGACTGTCTCCTTGCTATTAACTAGT ACAAGATTTGGGCTC 3' |
| (126)[2] | 2' | 5' GAGCCCAAATCTTGTACTAGTTAATAGCAAGGAGACAGT CATAATGAAATACCTATTGCCTACGGCAGCCGCTGGATT GTTATTAATCGCTGCCCAACCTGCCATGGCTGAGCTCGT GATGACCCACTCTCC 3' |

[1] Primes mRNA (sense strand) of heavy chain $C_H1$ region; antisense strand of light chain $V_L$ with dicistronic bridge in between heavy and light chains will be in the same relative orientation as given in the example.

[2] Primes antisense strand of heavy chain $C_H1$ regions; and sense strand of light chain $V_L$ region with dicistronic in between heavy and light chains will be in the same relative orientation as given in the example.

The resultant single step internally primed dicistronic DNA molecule can then be ligated into modified ImmunoZAP H for construction of an expression vector as described in Example 30.

27. Preparation of Lambda Zap II Expression Vector

The vector Lambda Zap™ II (Stratagene) is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the form of a phagemid (Bluscript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, (1988), by replacing the Lambda S gene contained in a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard protocols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, NY, 1987.

28. Preparation of $V_H$-Expression Vectors, ImmunoZAP H and Modified ImmunoZAP H, Construction

A. ImmunoZAP H

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is the most efficient method of reintroducing DNA into host cells. Second, it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. The alternative, plasmid cloning vectors, are only advantageous in the analysis of

clones after they have been identified. This advantage is not lost in the present system because of the use of lambda Zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be excised.

To express the plurality of $V_H$-coding DNA homologs in an E. coli host cell, a vector was constructed that placed the $V_H$-coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the $V_H$-coding DNA homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 25A. The individual single-stranded polynucleotides ($N_1$-$N_{12}$) are shown in Table 13 below.

Table 13

| Seq. Id. No. | | |
|---|---|---|
| (91) | N1) | 5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3' |
| (92) | N2) | 5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3' |
| (93) | N3) | 5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3' |
| (94) | N4) | 5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTTAATAG 3' |
| (95) | N5) | 5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3' |
| (96) | N6) | 5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3' |
| (97) | N7) | 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3' |
| (98) | N8) | 5' GTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGC 3' |
| (99) | N9-4) | 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3' |
| (100) | N11) | 5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3' |
| (101) | N12) | 5' TCGACGAATTCTATTAAGAACCGTAGTC 3' |

(102) N10-5) 5' CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

Polynucleotide N2, N3, N9-4', N11, N10-5', N6, N7 and N8 were kinased by adding 1 $\mu$l of each polynucleotide (0.1 ug/$\mu$l) and 20 units of $T_4$ polynucleotide kinase to a solution containing 70 mM Tris-HCl at pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 10 mM beta mercaptoethanol, 500 ug/ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides, 20 ng, of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 25A. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 $\mu$l of the above reaction to a solution containing 50 mM Tris-HCl, pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 $\mu$l of the above reaction, 4 $\mu$l of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a lambda Zap II vector prepared in Example 27 that had been previously digested with the restriction enzymes NotI and XhoI. The ligation mixture was packaged according to the manufacturer's instructions using Gigapack II Gold packing extract (Stratagene). The pack-

aged ligation mixture was plated on XL1-blue cells (Stratagene). Individual Lambda Zap II plaques were cored and the inserted excised according to the in vivo excision protocol provided by the manufacturer (Stratagene). This in vivo excision protocol converts the cloned insert from the Lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described in by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV Reverse Transcriptase $^{35}$S-ATP sequencing kit (Stratagene). The sequence of the resulting $V_H$ expression vector is shown in Figure 25A and Figure 26.

B. Modified ImmunoZAP H

To create a fusion-PCR library from hybridoma RNA for expressing the plurality of $V_H$-coding DNA homologs in an E. coli host cell, a vector based on the ImmunoZAP H vector described above was constructed. The procedure for constructing the vector was performed as described above with the following modifications: elimination of the SacI site between the $T_3$ polymerase and NotI sites and changing the nucleotide base residue sequence from AAA to CAG which resulted in an amino acid residue change from lysine to glutamine as shown in Figures 24A and 24B.

The individual single-stranded polynucleotides ($N_1$, $N_4$, $N_6$ and $N_7$), which were modified from their counterparts listed in Table 14, are listed in Table 14 below.

Table 14

| Seq. Id. No. | | |
|---|---|---|
| (127) | N1) | 5' AGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3' |
| (128) | N2) | 5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3' |
| (129) | N3) | 5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3' |
| (130) | N4) | 5' AGGTGCAGCTGCTCGAGAATTCTAGACTAGGTTAATAG 3' |

(131) N5) 5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
(132) N6) 5' CAGCTGCACCTGGGCCATGGCTGGTTGGG 3'
(133) N7) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(134) N8) 5' CTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGCGGCCGC 3'
(135) N9-4) 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
(136) N11) 5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
(137) N12) 5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
(138) N10-5) 5' CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

The modified ImmunoZAP H vector was created to eliminate an unnecessary SacI site in the ImmunoZAP H vector, (Example 28a, when the heavy and light chain vectors were combined. The modifications also improved the efficiency of secretion of positively changed amino acids in the amino terminus of the expressed protein. Inouye et al., Proc. Natl. Acad. Sci. USA, 85:7685-7689 (1988).

29. Preparation of $V_L$ Expression Vector ImmunoZAP L Construction

To express the plurality of $V_L$ coding polynucleotides in an E. coli host cell, a vector was constructed that placed the $V_L$ coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the periplasmic space and also provided a polynucleotide that coded for a spacer protein between the $V_L$ polynucleotide. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 25B. The individual single-stranded polynucleotides ($N_1$-$N_8$) are shown in Table 13 above.

Polynucleotides N2, N3, N4, N6, N7 and N8 were kinased by adding 1 $\mu$l of each polynucleotide and 20 units

of $T_4$ polynucleotide kinase to a solution containing 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DDT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all of the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40 $\mu$l of the above reaction to a solution containing 50 $\mu$l Tris-HCl, pH 7.5, 7 mM $MgCl_2$ 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 $\mu$l of the above reaction, 4 $\mu$l of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a Lambda Zap II vector prepared in Example 27 that had been previously digested with the restriction enzymes NotI and XhoI. The ligation mixture was packaged according to the manufacturer's instructions using Gigapack II Gold packing extract and the packaged ligation mixture was plated on XL1-Blue cells as described in Example 28A. Individual lambda Zap II plaques were cored and the inserts excised according to the <u>in vovo</u> excision protocol as described in Example 28A. This <u>in vivo</u>

excision protocol converts the cloned insert from the Lambda Zap II vector into a phagemid vector to allow easy manipulation and sequencing and also produces the phagemid version of the $V_L$ expression vectors. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase $^{35}$S-dATP sequencing kit (Stratagene). The sequence of the resultin $V_L$ expression vector is shown in Figure 25B and Figure 27).

The $V_L$ expression vector used to construct the $V_L$ library was the phagemid produced to allow the DNA of the $V_L$ expression vector to be determined. The phagemid was produced, as detailed above, by the in vivo excision process from the Lambda Zap $V_L$ expression vector (Figure 27).

30. Construction of $V_{HL}$-Expression Vectors and Library

A. Ligation of Dicistronic DNA Molecules with Modified ImmunoZAP H

In preparation for cloning a library enriched in $V_H$-$V_L$-coding ($V_{HL}$) dicistronic DNA molecules, PCR amplified products (human or mouse) prepared in Examples 24, 25, and 26 (50 mM NaCl, 25 mM Tris-HCl, pH 7.7, 10 mM $MgCl_2$, 10 mM $\beta$-mercaptoethanol, 100 ug/ml BSA, at 37°C were digested with restriction enzymes XhoI and XbaI at a concentration of 60 units of enzyme per ug of DNA, and purified on a 1% agarose gel. After gel electrophoresis of the digested PCR amplified dicistronic DNA molecules, the region of the gel containing the DNA fragments of approximately 1360 base pairs in size was excised, purified using Gene-Clean (BIO 101, La Jolla, California), ethanol precipitated and resuspended in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA to a final concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 4°C to 1 ug of modified ImmunoZAP H vector, prepared in Example 28b, (Stratagene) previously digested with XhoI and XbaI. A

portion of the ligation mixture (1 ul) was packaged for 2 hours at room temperature using Gigapack Gold packaging extract (Stratagene) and the packaged material was plated on a permissive E. coli (strain XL1-blue) lawn to generate plaques. The library was determined to consist of predominantly $V_{HL}$ with less than 5% non-recombinant background.

B. Screening of Antibody-Producing Plaques

(i) Human

To screen for expression of $V_{HL}$ dicistronic molecules, E. coli were infected to yield approximately 100 plaques per plate. Replica filter lifts of the plaques on an agar plate were produced by overlaying a nitrocellulose filter that had been soaked in 10 mM isopropyl beta-dithiogalactopyranoside on each plate with transfer for 15 hours at 23°C. For detection of $V_{HL}$ antibody fragment expression, the filters were screened with rabbit anti-human heavy and light chain antibodies followed by goat anti-rabbit antibody coupled to alkaline phsophatase (Cappel Laboratories, Malver, Pennsylvania). The detection of immunoreactive product confirmed the presence and expression of $V_{HL}$ antibody fragments.

To identify human DNA clones expressing antibody that bound TT, plaques were plated and proteins expressed as described above. Replica filters were incubated with 0.2 nN $^{125}$I-tetanus toxoid and washed. Positive plaques were identified by autoradiography and isolated. The frequency of positive clones in the library was equivalent to (number of positive clones)/[number of plaques screened) X (fraction of plaques expressing $V_{HL}$). Concentrated non-adsorbed tetanus toxoid was iodinated with sodium iodide $^{125}$I (ICN, Irvine, California) by the Choramine-T method as described in Botton et al., Biochem. I., 133:529-539 (1973) and available in a kit (Iodo-Beads, Pierce, Rockford, Illinois).

Human DNA clones were re-plated at approximately 100 phage per plaque side by side with the parental phage that

were used as templates for PCR amplification and screened in the primary antigen binding screen. The results of the screening procedure are seen in Figure 28. Similar signals between the parental clones and the $V_{HL}$ dicistronic DNA molecules demonstrated that the sequence differences introduced with the $C_H1'$ and $V_L$ primers did not adversely affect gene expression. Also, it should be noted in Figure 28 that a random parental clone that did not react with tetanus toxoid, 7G1, was unreactive before and after the PCR dicistronic fusion, as was the control ImmunoZAP H vector (IZ H).

(ii) Mouse

Mouse antibody-producing plaques prepared in Example 27 were screened for antibody expression with rabbit anti-mouse heavy and light chain antibody (Cappel Laboratories) as described above.

31. Characterization of Cloned Dicistronic $V_{HL}$ Repertoire in Expression Library

A. Verification of Presence and Size of Cloned Dicistronic $V_{HL}$ Repertoire

Bacteriophage from purified reactive plaques prepared in Example 30B were converted to the plasmid format by in vivo excision with R408 helper phage according to manufacturer's protocol (Stratagene) and also described in Short et al., Nucl. Acids Res., 16:7583-7600 (1988). In the in vivo excision protocol, the cloned insert from the ImmunoZAP H vector was converted into a phagemid vector to allow easy manipulation and sequencing. Briefly, phage plaques were cored from the agar plates and transferred to sterile microfuge tubes containing 500 ul of a buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$, and 0.01% (w/v) gelatin and 20 ul of chloroform.

For excisions, 200 ul of the phage stock, 200 ul of XL1-Blue cells ($A_{600}$ = 1.00) and 1 ul of R408 helper phage ($1 \times 10^{10}$ plaque forming units Opfu)/ml) were incubated at

37°C for 15 minutes. AFter a 4 hour incubation in Luria-Bertani (LB) broth and heating at 70°C for 20 minutes to heat kill the XL1-blue cells, the phagemids were re-infected into XL1-Blue cells and plated onto LB plates containing ampicillin. Double stranded DNA was prepared from the phagemid containing cells according to the methods described by Holmes et al., Anal. Biochem., 114:193, (1981). Clones were first screened for DNA inserts by restriction digests with Xho1 and Xba1. The detection of 1390 base pair fragment on an agarose gel confirmed the presence of a $V_{HL}$ dicistronic molecule insert.

B. Sequencing of Plasmids from Expression Library

Clones containing the putative $V_{HL}$ insert were sequenced using reverse transcriptase according to the general method described by Sanger et al., Proc. Natl. Acad. Sci., USA, 74:5463-5467, (1977) and the specific modifications of this method provided in the manufacturer's instructions in the AMV reverse transcriptase $^{35}S$-dATP sequencing kit (Stratagene).

Nucleotide sequence analysis of several fusion clones indicated that the sequence of the fusion region was identical to that shown in Figure 22, proving that the clones were actually generated through a fusion PCR intermediate.

C. Advantages of Fusion-PCR to Produce Dicistronic DNA Molecules

PCR amplification can, therefore, be used to fuse sequences responsible for encoding subunits of a heterodimeric protein together into a single DNA fragment that can then direct the expression of both subunits from one expression vector. In the case of antibodies, if the source of nucleic acid template comes from hybridoma mRNA, there is only one heavy and light chain sequence to choose from, and thus the heavy:light pair is a "natural" pair.

However, if spleen, peripheral blood B-cell, or other lymphocyte mRNA is used as the source of template, the PCR fusion reaction to form a dicistronic DNA molecule can randomly pair heavy and light chains from different cells, producing a combinatorial library. In such a library, only a small fraction of the clones contain the original heavy and light chain pairs. This may not be a problem if the desired natural pair is well represented in the original B-cell population, as is the case with hyperimmunized donors. However, if one wishes to find a naturally occurring rare specificity in a combinatorial library, one may have to screen a large number of clones.

The fusion method presented here may offer a solution to the random combinatorial problem. If one begins with a very dilute population of B-cells (possibly in a medium that limits diffusion), it may be possible for the dicistronic event to occur between naturally paired heavy and light chain sequences before significant mixing between B-cell RNA occurs. Thus, the fused heavy and light chain sequences would be the original pairs, and the resulting library would express predominantly the naturally occurring antibody specificities. Such a library would be highly preferable when rare natural specificities are sought.

Another advantage to this method is that only one vector and one cloning step are necessary. This saves a substantial amount of time, resources, and effort. Moreover, the ease of the single PCR reaction greatly simplified the process of going from B-cell RNA to an E. coli library, making this approach a noteworthy alternative to standard hybridoma technology.

The foregoing is intended as illustrative of the present invention but not limiting. Numerous variations and modifications can be effected without departing from the true spirit and scope of the invention.

Claims:

1. A method of producing a nucleic acid vector encoding two or more desired genes, each from a family of genes, said genes being capable of together producing a characteristic that can be used to identify the vector encoding said desired genes from other vectors encoding other combinations of genes from said families of genes, which method comprises:

a) randomly inserting into vectors one member from a first family of genes and one member from one or more other families of genes so that a population of vectors are created wherein each vector may contain one of the genes from said first gene family and one of the genes from each of said other gene families;

b) identifying within said population of vectors a vector capable of detectably producing a desired characteristic resulting from the inclusion of one gene from said first gene family and one gene from each of said other gene families, and using said characteristic to distinguish the vector from other vectors within the population containing undesired combinations of gene members from said gene families.

2. The method of claim 1 wherein said genes are inserted into a DNA vector at one or more integration sites, which method further comprises:

a) preparing said vectors with one or more site- or region-specific recombination sequences;

b) permitting, in the presence of one or more reagents facilitating said site- or region-specific recombination, a member of said first family of genes to combine in a vector with a member of said second family of genes.

3. The method of claim 2 wherein said site- or region-specific recombination site is recognized and acted on by flp recombinase.

4. The method of claim 2 wherein said site- or region-specific recombination site is recognized and acted on by cre recombinase.

5. The method of claim 2 wherein said site- or region-specific recombination site is recognized and acted on by lambda integrase recombinase.

6. The method of claim 2 wherein at least one of the vectors contains a sequence capable of being recognized and acted on by transposase.

7. The method in claim 1 where said genes are inserted into a DNA vector at one or more integration sites, which method further comprises:

a) cleaving said vector with one or more site-specific integration reagents;

b) preparing the ends of genes from said first family of genes so that one end will ligate with an end of the vector cleaved by a first reagent and the other with an end of the vector cleaved by a second reagent;

c) preparing the ends of said genes from said other gene families so that one end will ligate with an end of the vector cleaved by a third reagent and the other with an end of the vector cleaved by a fourth reagent;

d) preparing at least one double stranded DNA linker fragment having one end ligatable to one end of said genes from said first family of genes and the other end ligatable to one end of genes from said other family of genes;

e) mixing said vector, genes, and said linker fragment or fragments together in a ligation mix and ligating the components.

8. The method of claim 7 wherein said reagents are the same.

9. the method of claim 8 wherein said reagents are different.

10. The method of claim 1, wherein said combination of genes is accomplished in vivo.

11. A method of producing a host cell expressing two or more desired genes, each from a family of genes, said genes being capable of together producing a characteristic that can be used to identify the host cell expressing said desired genes from other host cells expressing other combinations of genes from said families of genes, which method comprises:

a) randomly introducing into host cells one member from a first family of genes and one member from one or more other families of genes so that a population of host cells are created wherein each host cell may contain one of the genes from said first gene family and one of the genes from each of said other gene families;

b) identifying within said population of host cells a host cell capable of detectably exhibiting a desired characteristic resulting from the inclusion of one gene from said first gene family and one gene from each of said other gene families, and using said characteristic to distinguish the host cell from other host cells within the population containing undesired combinations of gene members from said gene families.

12. The method of claim 11 wherein said vectors are lambda bacteriophage vectors and the host cells are E. Coli.

13. A method of producing a nucleic acid vector encoding two or more genes belonging to families of genes, being capable of producing a characteristic that can be used to identify the vector encoding said genes from other

vectors encoding other members of the families of genes which method comprises:

a) isolating a first population of vectors for which each member of said population may contain one member of a family of genes;

b) inserting one member of a second family of genes into each of the vectors so that a population of vectors are created where each vector may contain one of the genes from said first family and one of the genes from said second family;

c) identifying within said population of vectors a vector capable of producing a characteristic resulting from the inclusion of one gene from said first gene family and one gene from said second gene family, and using said characteristic to distinguish the vector from other vectors within the population containing other members of the gene families.

14. A method of producing a nucleic acid vector encoding two or more genes belonging to families of genes, said genes being capable of producing a characteristic that can be used to identify the vector encoding said genes from other vectors encoding other members of the families of genes, which method comprises:

a) isolating a first population of vectors, for which each member of said population may contain one member of a first family of genes and a nucleic acid site or region at which the population of vectors can be combined with a second population of vectors;

b) isolating a second population of vectors, for which each member of said population may contain one member of a second family of genes and a nucleic acid site or region at which the second population of vectors can be recombined with said first population of vectors so that one member of the first family of genes and one member of the second family of genes may be combined and expressed

in each member of a diverse population of recombined vectors;

c) recombining populations of said first and second vectors and at said nucleic acid site or region thereby creating a diverse population of recombinant vectors each of which may express one member of the first family of genes and one member of the second family of genes;

d) identifying within said population of recombinant vectors a vector capable of producing a characteristic resulting from the inclusion of one gene from each of said gene families.

15. The method of claim 14 wherein said nucleic acid site is cleaved with site-specific reagent, which method further comprises:

a) cleaving said first vector population with said reagent;

b) cleaving said second vector population with said reagent;

c) mixing both vector populations together in a ligation mix and ligating the two populations.

16. The method of claim 14 wherein said nucleic acid region is a homologous region capable of undergoing homologous recombination, which method further comprises inserting one or more members of said first and second populations into a single host capable of carrying out homologous recombination and allowing such homologous recombination to occur.

17. The method of claim 14 wherein said nucleic acid site is a target site for site-specific recombination, which method further comprises inserting one or more members of said vector populations into a single host capable of carrying out site-specific recombination at said nucleic acid site and allowing said site-specific recombination to occur.

18. The method of claim 17 wherein said target site for site-specific recombination is of the family of sites selected from flp, lox, and gamma-delta.

19. The method of any of claims 1, 11, 13 or 14 wherein said vectors are plasmid or cosmid vectors.

20. The method of any of claims 1, 11, 13 or 14 wherein said vectors are phage vectors.

21. The method of any of claims 1, 13 or 14 wherein said vectors are lambda bacteriophage vectors.

22. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the interaction of sequence-specific nucleic acids with genes from said first and second families of genes.

23. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the hybridization of nucleic acid probes with genes from said first and second of families of genes.

24. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the expression of one or both of genes from said gene families as an RNA molecule.

25. The method of claim 14 wherein the identification of a particular vector within the recombinant vector population involves the expression of one or both of genes from said gene families as an identifiable protein molecule.

26. The method of claim 25 wherein the protein molecule(s) contains a binding site for another molecule.

27. The method of claim 26 wherein the protein molecule(s) contains an epitope recognized by an antibody.

28. The method of claim 27 wherein the protein molecule(s) contains an immune molecule binding site for an epitope.

29. The method of claims 14 wherein both genes express an RNA and/or polypeptide and said RNAs and/or polypeptides physically interact within a host to create said characteristic.

30. The method of claim 29 wherein both genes express polypeptides that physically interact to form a neo-epitope recognized by an immune molecule.

31. The method of claim 29 wherein both genes express polypeptides that physically interact to form a binding site for another molecule.

32. The method of claim 31 wherein the polypeptides are derived from antibody genes such that the interaction of both polypeptides forms an antigen binding site.

33. The method of any of claims 1, 11, 13 or 14 wherein the vectors contain a single promoter that expresses the genes from said gene families.

34. The method of any of claims 1, 11, 13 or 14 wherein said genes from said gene families are each expressed from their own promoter.

35. The method of claim 11 wherein the host is a mammalian cell.

36. The method of claim 11 wherein the host is a eukaryotic cell.

37. The method of claim 11 wherein the host is a prokaryotic cell.

38. The method of any of claims 1, 11, 13 or 14 wherein there are more than two gene families and the vectors produced contain a random assortment of one member of each gene family needed to create said characteristic.

39. HCFLP.

40. LCFLP.

41. A method of producing a biological agent having a desired phenotype wherein said phenotype results from expression of a particular combined nucleotide sequence and wherein said phenotype can be used to identify the biological agent having the particular combined nucleotide sequence which comprises:

(a) bringing together a first population of nucleotide sequences with one or more other populations of nucleotide sequences to produce combined nucleotide sequences wherein each separate combined nucleotide sequence comprises one member of each population of nucleotide sequences;

(b) expressing said combined nucleotide sequences in biological agents; and

(c) identifying those biological agents which express said desired phenotype.

42. A method according to claim 41 wherein said phenotype can be used to distinguish the biological agent from bioological agents having other combined nucleotide sequences further comprising using said phenotype to distinguish those biological agents expressing the

particular combined nucleotide sequence from biological agents having other combined nucleotide sequences.

43. A method according to claim 41 wherein said biological agent is a cell.

44. A method according to claim 41 wherein said biological agent is nucleic acid vector.

45. A method according to claim 41 wherein said biological agent is a bacteriophage or virus.

46. A method according to claim 41 wherein said phenotype results from expression of a hybrid polypeptide which is encoded by the particular combined nucleotide sequence and is encoded at least in part by one nucleotide sequence from each population of nucleotide sequences which was brought together.

47. A method according to claim 41 wherein said phenotype results from expression of a plurality of polypeptides wherein a polypeptide is encoded at least in part by one nucleotide sequence from each separate population of nucelotide sequences which was brought together.

48. A method according to claim 41 wherein two populations of nucleotide sequences are combined.

49. A method according to claim 47 wherein said phenotype results from expression of a heterodimeric polypeptide wherein one subunit of said dimer is encoded at least in part by the nucleotide sequence from the first population of nucleotide sequences and the other subunit of said dimer is encoded at least in part by the nucleic sequence from the second population of nucleotide sequences.

50. A method according to claim 48 wherein said phenotype results from expression of a first polypeptide encoded at least in part by the nucleotide sequence from the first population of nucleotide sequences and of a second polypeptide encoded at least in part by the nucleotide sequence from the second population of nucleotide sequences.

51. A method according to claim 48 wherein said phenotype results from expression of an RNA molecule encoded at least in part by the nucleotide sequence from the first population of nucleotide sequences and a second RNA molecule encoded at least in part by the nucleotide sequence from the second population of nucleotide sequences.

52. A method according to claim 48 wherein said phenotype results from synthesis of an RNA molecule encoded at least in part by the nucleotide sequence from the first population of nucleic acid sequences and by the nucleic acid sequence from the second population of nucleic acid sequences.

53. A method according to claim 48 wherein the first and second populations of nucleotide sequences are combined by co-infection or co-transformation of host cells.

54. A method according to claim 48 wherein members from said first and second populations of nucleotide sequences are combined randomly to give combined nucleotide sequences.

55. A method according to claim 41 wherein the combining of said populations of nucleotide sequences gives a combined nucleotide sequence which was not previously expressed in said biological agent.

56. A method according to claim 41 wherein said desired phenotype comprises a phenotype which was not previously expressed in a population of such biological agents.

57. A method according to claim 41 wherein said first population of nucleotide sequences comprises non-identical nucleotide sequences.

58. A method according to claim 41 wherein each population of nucleotide sequences comprises non-identical nucleotide sequences.

59. A method of producing a nucleic acid vector encoding a preselected combined nucleotide sequence which comprises two or more preselected nucleotide sequences, each independently selected from a population of nucleotide sequences, said combined nucleotide sequence being capable of producing a characteristic that can be used to identify the vector encoding said preselected combined nucleic sequence comprises

(a) bringing together a member nucleotide sequence from each population of nucleotide sequences to give a population of combined nucleotide sequences wherein each combined nucleotide sequence comprises a nucleotide sequence from each population;

(b) inserting into vector a member of the population combined nucleotide sequences so that a population of vectors is created wherein each vector may contain a combined nucleic acid sequence;

(c) identifying within said population of vectors, a vector capable of detectably producing a desired characteristic resulting from inclusion of the preselected combined nucleic acid sequence.

60. A method according to claim 59 wherein said characteristic can be used to distinguish the vector

encoding the preselected combined nucleotide sequence from other vectors encoding other combinations of nucleotide sequences further comprising using said characteristic to distinguish the vector from other different vectors within the population having unselected combined nucleotide sequences.

61. A method according to claim 60 wherein said nucleotide sequences are combined randomly.

62. A method according to claim 61 wherein said combined nucleotide sequences are produced using fusion polynucleotide amplification.

63. A method according to claim 59 wherein said combined nucleotide sequences are produced using fusion polynucleotide amplification.

64. A method according to claim 1 wherein a dicistronic or multicistronic DNA sequence which comprises one member from the first family of genes and one member from one or more than families of genes which comprises a random combination of said members of said families of genes is synthesized using fusion polynucleotide amplification and inserted into vectors.

65. A method for producing a biological agent having a desired novel phenotype wherein said phenotype results from expression of a particular combined nucleotide sequence and wherein said phenotype can be used to identify the biological agent having 'the particular combined nucleotide sequence; which comprises:

(a) replicating at least portions of at least two parent nucleotide sequences under conditions that allow mutations to occur in either nucleotide sequence to generate a population of diverse replicas of each parent nucleotide sequence;

(b) randomly bringing together the populations of diverse replicas to produce combined nucleotide sequences wherein each combined nucleotide sequence comprises one member of each population of diverse replicas;

(c) expressing said combined nucleotide sequences in biological agents; and

(d) identifying those biological agents which express said desired phenotype.

66. A method according to claim 65 wherein said desired phenotype is distinguishable from phenotypes expressed by said parent nucleotide sequences.

67. A method according to claim 66 wherein said phenotype can be used to distinguish it from biological agents having other combined nucleotide sequences using said phenotype to distinguish those biological agents expressing the particular combined nucleotide sequence from biological agents having other combined nucleotide sequences.

68. A method according to claim 65 wherein the parent nucleotide sequences comprise a single DNA molecule and are replicated together; further comprising separating the populations of diverse replicas of each parent nucleotide sequence prior to bringing together step (b).

69. A method according to claim 68 which comprises replicating two parent nucleotide sequences.

70. A method according to claim 65 wherein the parent nucleotide sequences are separately replicated.

71. A method according to claim 70 which comprises replicating two parent nucleotide sequences.

72. A method according to 71 wherein a first parent nucleotide sequence is replicated in one population of cells and a second parent nucleotide sequence is replicated in a second population of cells and said cell populations are mixed and fused to generate cells which express combined nucleotide sequences.

73. A method according to claim 72 wherein said first parent nucleotide sequences codes for a selected $V_L$ and said second parent nucleotide sequences codes for a selected $V_H$, said cells are E. coli; and said combined nucleotide sequences express a Fab.

74. A method for producing a biological agent having a desired phenotype wherein said phenotype results from expression of a particular combined nucleotide sequence and wherein said phenotype can be used to identify the biological agent having the particular combined nucleotide sequence which comprises:

(a) replicating parent populations of nucleic acid sequences to generate a population of diverse replicas of each parent population:

(b) randomly bringing together the populations of diverse replicas to produce combined nucleotide sequences wherein each combined nucleotide sequence comprises one member of each population of diverse replicas;

(c) expressing said combined nucleotide sequences in biological agents; and

(d) identifying those biological agents which express said desired phenotype.

75. A method according to claim 74 wherein said desired phenotype is distinguishable from phenotypes expressed by said parent populations of nucleotide sequences.

76. A method according to claim 75 wherein said phenotype can be used to distinguish said biological agent from biological agents having other combined nucleotide sequences, further comprising using said phenotype to distinguish those biological agents expressing the particular combined nucleotide sequence from biological agents having other combined nucleotide sequences.

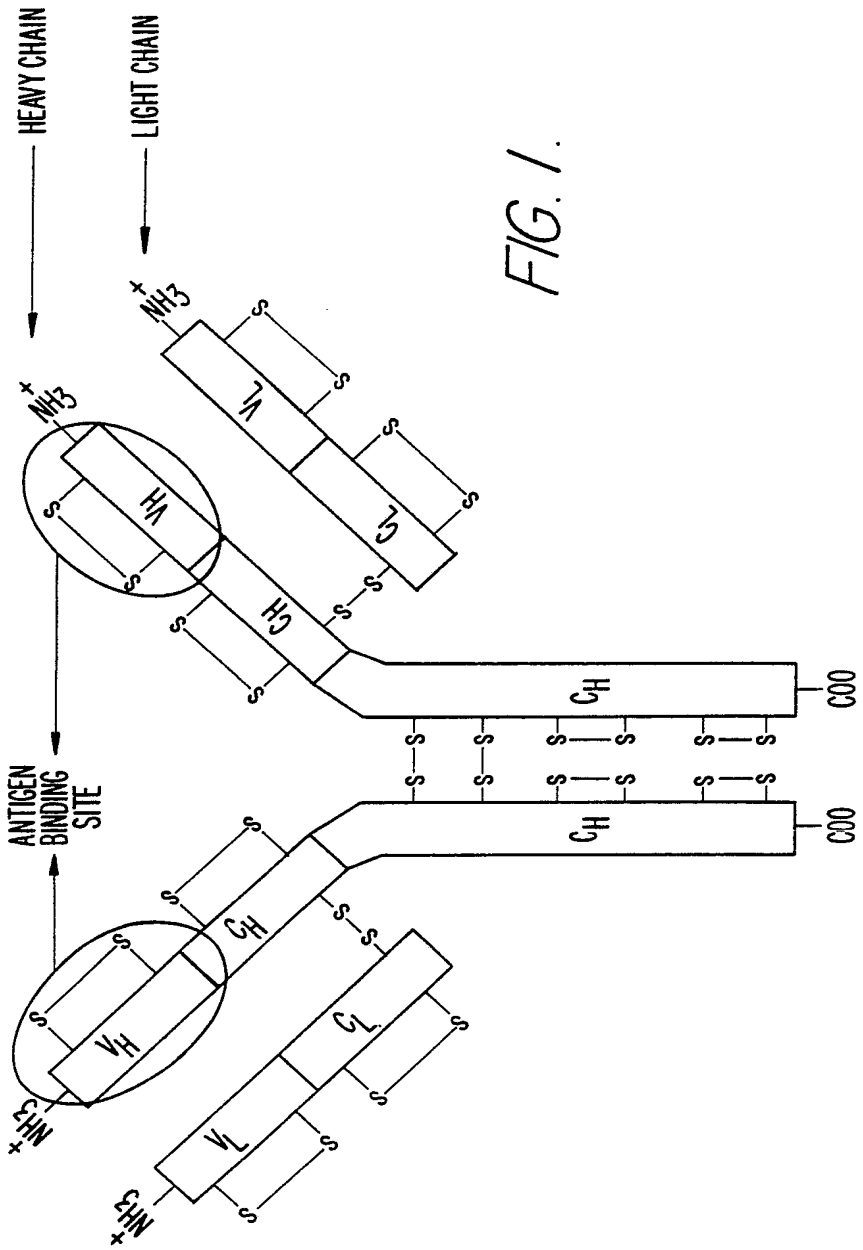



FIG. 1.

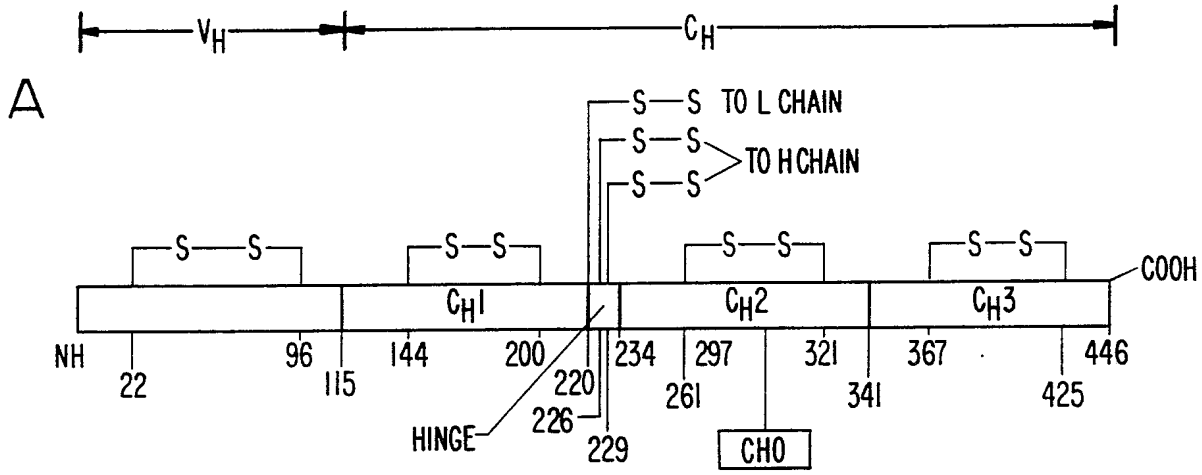

A
$V_H$
$C_H$
S—S TO L CHAIN
S—S
S—S
TO H CHAIN
S—S
S—S
S—S
S—S
COOH
$C_H1$
$C_H2$
$C_H3$
NH
22
96
115
144
200
220
226
229
234
261
297
321
341
367
425
446
HINGE
CHO

FIG. 2a.

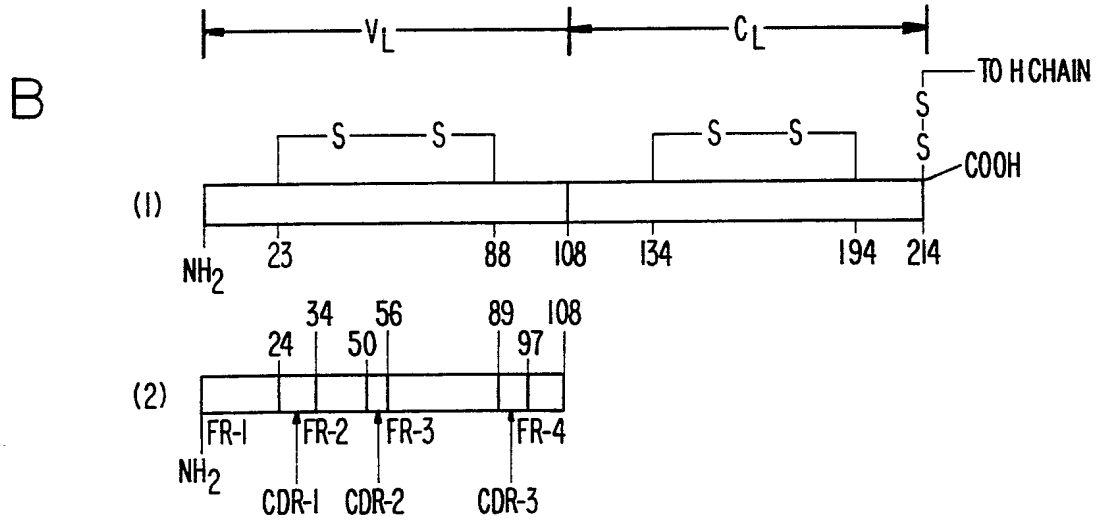

B
$V_L$
$C_L$
TO H CHAIN
S
S
S—S
S—S
COOH
(1)
$NH_2$
23
88
108
134
194
214
34
56
89
108
24
50
97
(2)
FR-1
FR-2
FR-3
FR-4
$NH_2$
CDR-1
CDR-2
CDR-3

FIG. 2b.

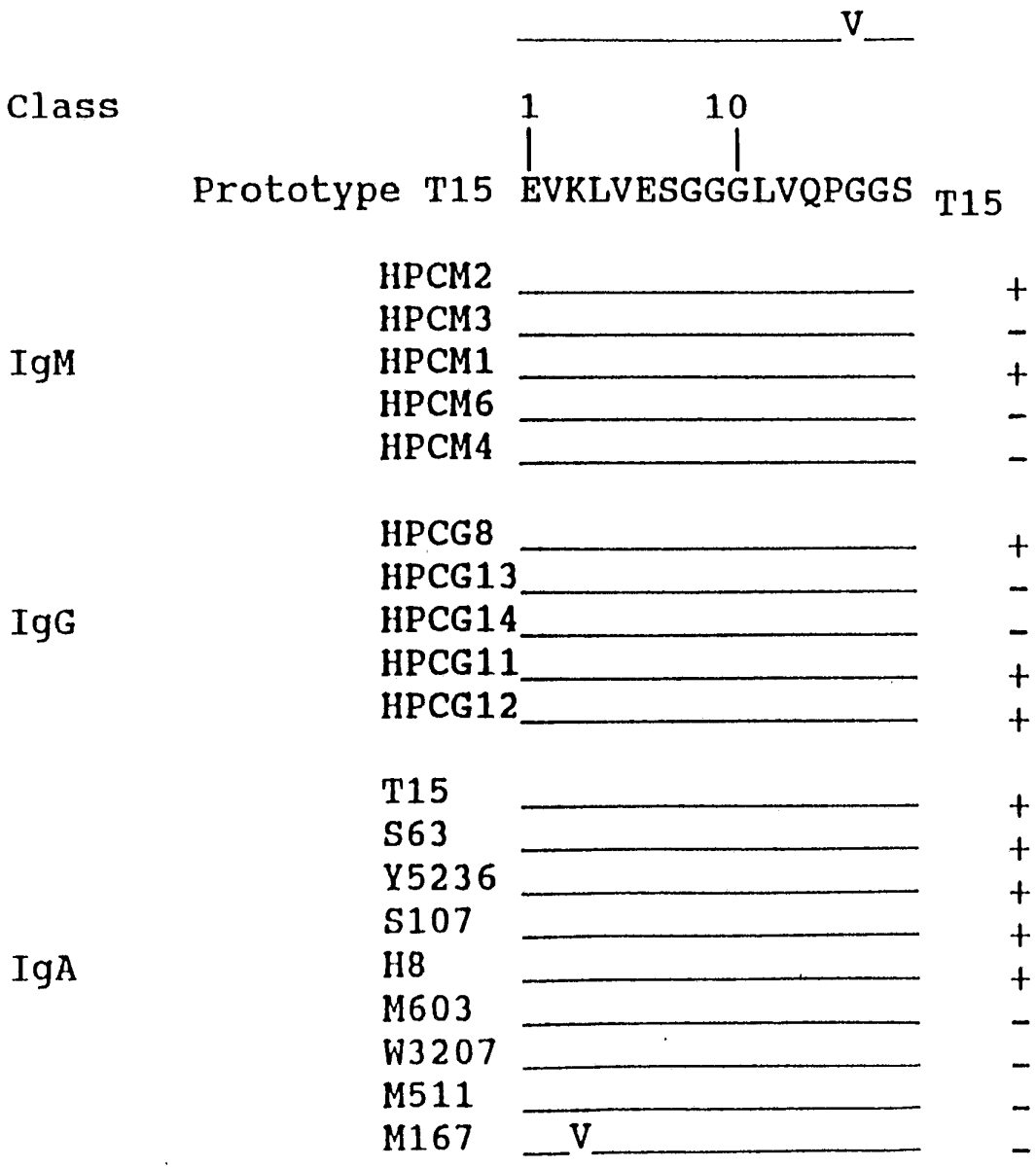
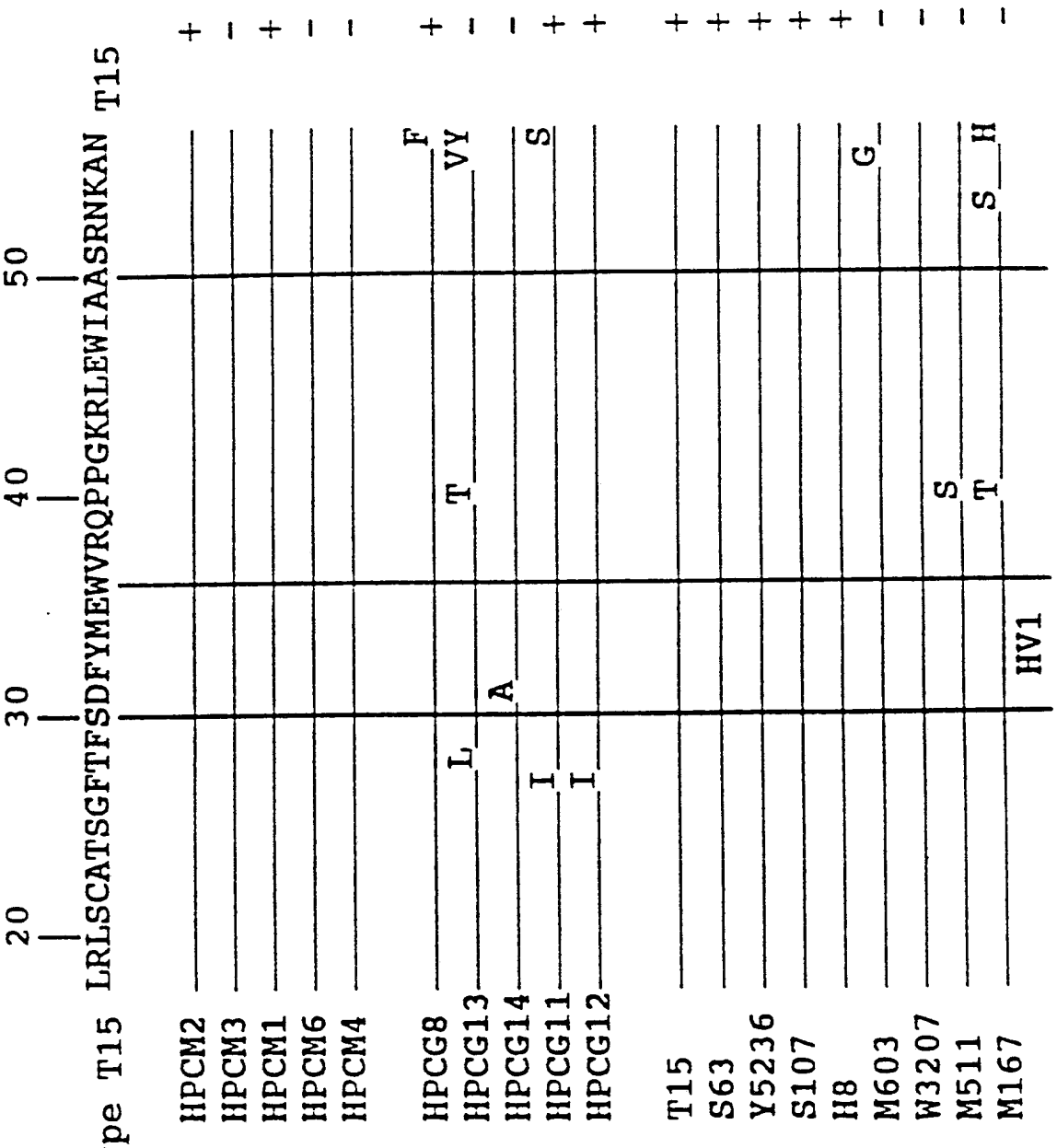



FIG. 3-1

FIG. 3-2


Class
Prototype T15
LRLSCATSGFTFSDFYMEWVRQPPGKRLEWIAASRNKAN T15
20 30 40 50
V
HV1
IgM
IgG
IgA
HPCM2
HPCM3
HPCM1
HPCM6
HPCM4
HPCG8
HPCG13
HPCG14
HPCG11
HPCG12
T15
S63
Y5236
S107
H8
M603
W3207
M511
M167

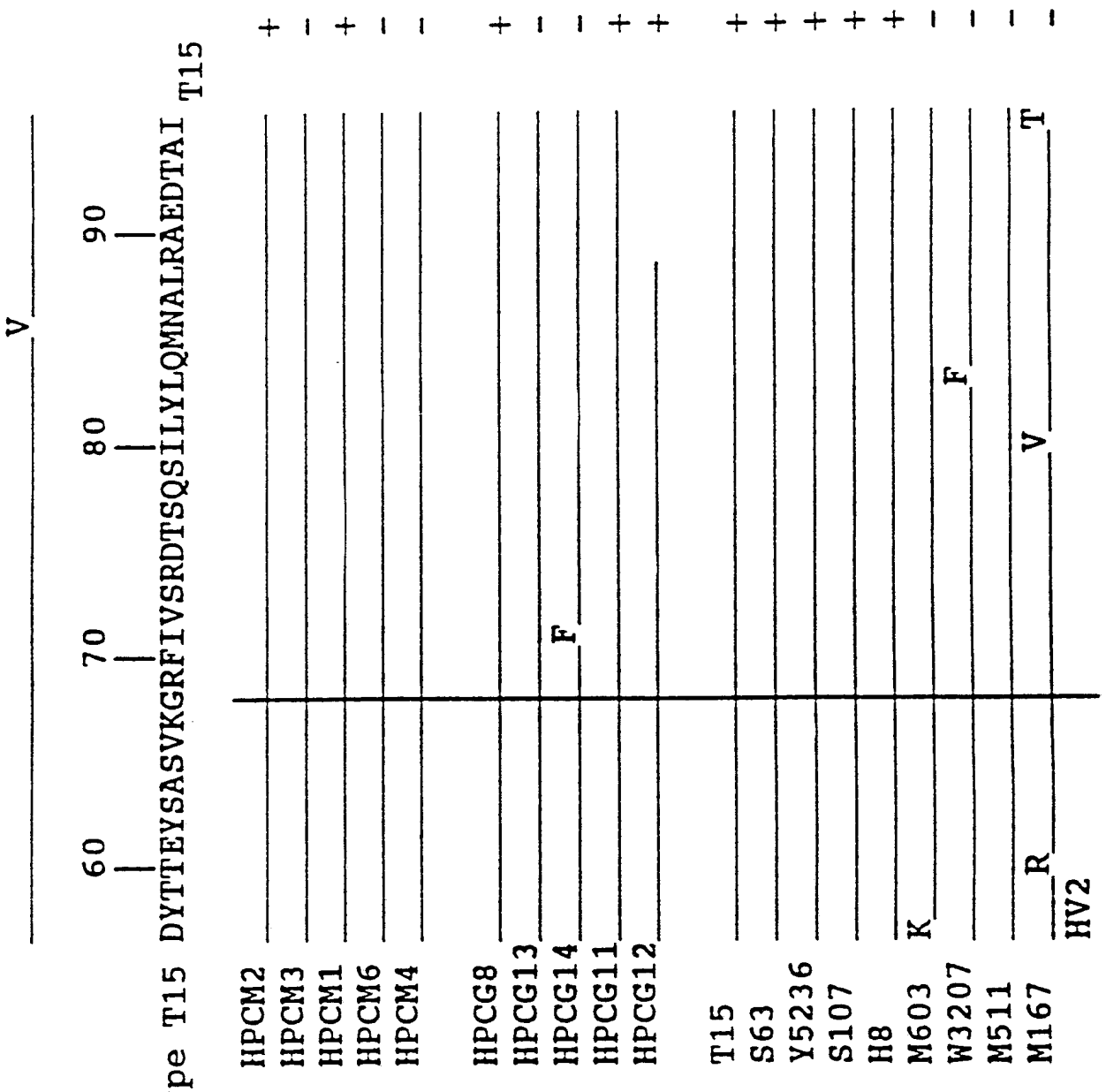

Class
Prototype T15 DYTTEYSASVKGRFIVSRDTSQSILYLQMNALRAEDTAI T15
60 70 80 90
V
IgM
HPCM2 +
HPCM3 -
HPCM1 +
HPCM6 -
HPCM4 -
IgG
HPCG8 +
HPCG13 -
HPCG14 - F
HPCG11 +
HPCG12 +
IgA
T15 +
S63 +
Y5236 +
S107 +
H8 +
M603 - K
W3207 - F
M511 - V T
M167 - R
HV2

FIG. 3-3

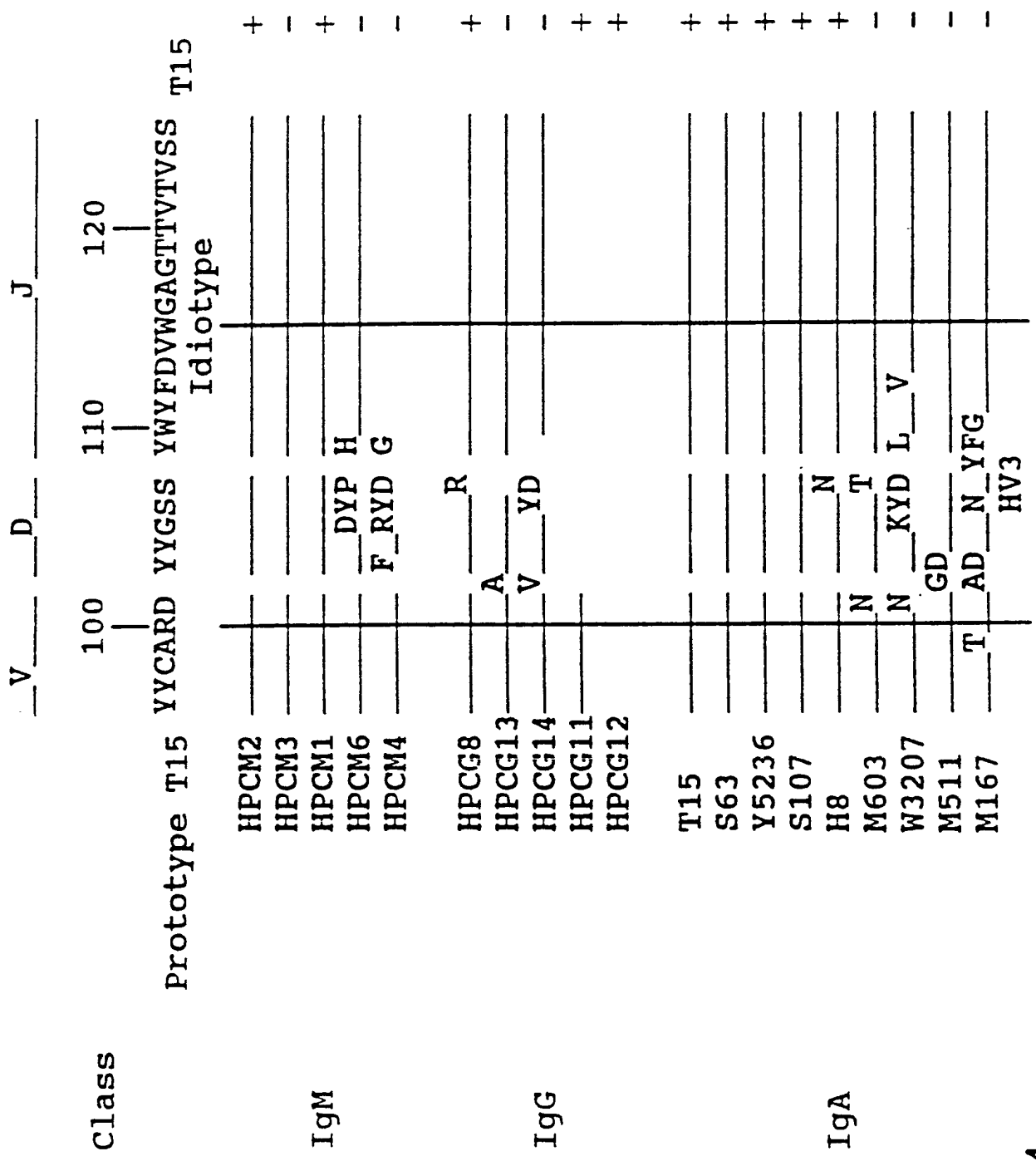

Class
Prototype T15 YYCARD YYGSS YWYFDVWGAGTTVTVSS T15
V D J
100 110 120
Idiotype
IgM
IgG
IgA
HPCM2
HPCM3
HPCM1 DYP H
HPCM6 F RYD G
HPCM4
HPCG8 R
HPCG13 A
HPCG14 V YD
HPCG11
HPCG12
T15
S63
Y5236
S107
H8 N
M603 N T
W3207 N KYD L V
M511 GD
M167 T AD N YFG
HV3
+ - + - - + - - - - + + + + + - - - -

FIG. 3-4

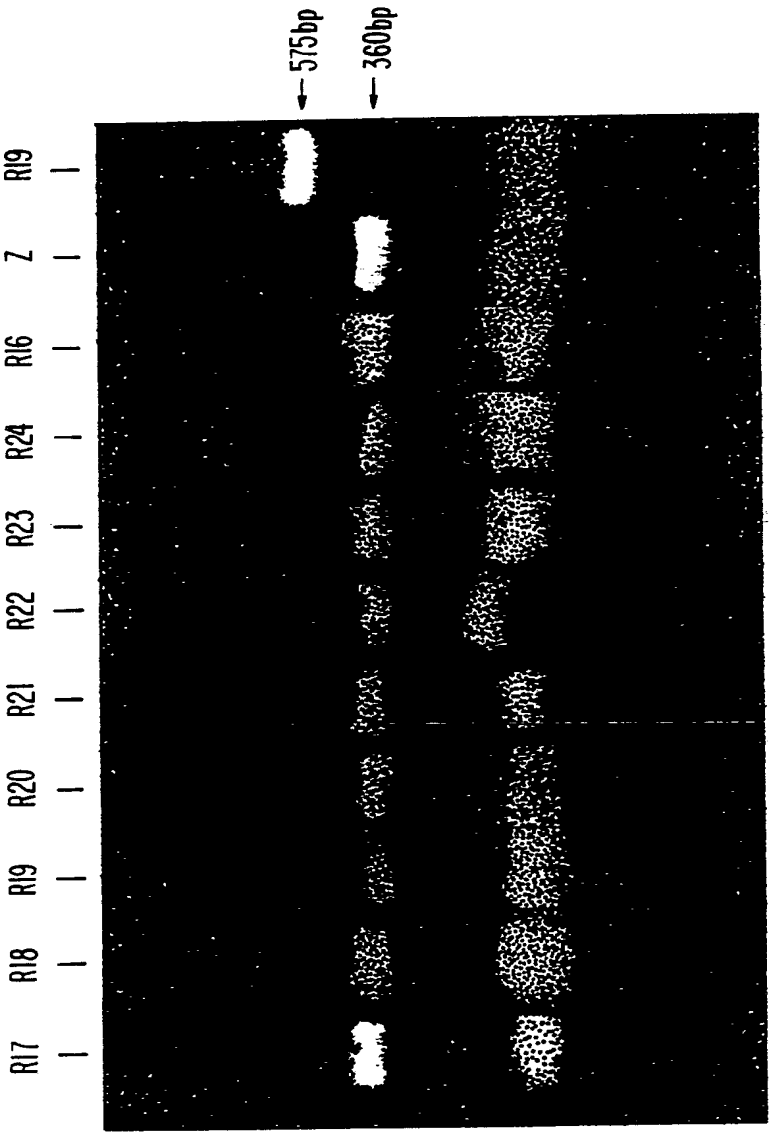



FIG. 4.

FIG. 5-1

Subclass I (A)
#L39

CTCGAGTCAGGACCTGGCCTCGTGAAACCTTCTCAGTCTCTGTCTCTC
ACCTGCTCTGTCACTGGCTACTCCATCACCAGTGCTTATTACTGGAAC
TGGATCCGGCAGTT

Subclass II (A)
#L11

CTCGAGTCTGGGCCTnAACTGGCAAAACCTGGGGCCTCAGTGAAGATG
TCCTGCAAGGCTTCTGGCCACACCTTGACTAGTTACTGGATACACTGG
GTAAAAnAGAGGCC

#L03

CTCGAGTCTGGACCTnAGCTGGTAAAGCCTGGGGTTCAGTGAAGATGT
CCTGCAAGGCTTCTGGATACACATTCACnAGCTATGTTATACACTGGG
TGAAGCAGAAGCCT

FIG. 5-2

#L32

CTCGAGTCTGGACCTGAACTGGTAAAGCCTGGGACTTCAGTGAAGATG
TCCTGCAAGGCTTCTGGATACACATTCACCAGCTATGTTATGCGCTGG
GTGAAGCAGAAGCC

Subclass II (B)
#L37

CTCGAGTCAGGGGCTGAACTGGTGAAGCCTGGGGTTTCAGTGAAGTTG
TCCTGCAAGGCTTCTGGCTACACCTTCACnAGCTACTATATGTACTGG
GTGAAGCAGAGGCC

#L06

CTCGAGTCTGGGGCTAAGCTGGTAAGGCCTGGAGCTTnAGTnAAGCTG
TCCTGnAGGGCTTCTGGCTACTCCTTCACnAGCTACTGGATGAACTGG
GTGAAGCAGAGGCC

FIG. 5-3

Subclass II (C)
#L33

CTCGAGTCTGGGGCTGAGCTGGTGAGGCCTGGAGCTTCAGTnAAGCTG
TCCTGCAAGGCCTCTCGTACTCCTTCACCAGCTCCTGATAACTGGGTG
AAGCAGAGGCCTGG

Subclass III (B)
#L36

CTCGAGTCAGGAGGTGGCCTGGTGCAGCCTGGAGGATCCCTGAAACTC
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAATTGG
GTCCGGCAGCTCCA

#L02

CTCGAGTCTGGAGGTGGCCTGGTGCAGCCTGGAGGATCCCTGAATCTC
CCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATAATGGATGAGTTGG
GTTCGGCAGGCTCC

FIG. 5-4

#L31

CTCGAGTCTGGAGGTGGCCTGGTGCAGCCTGGAGGATCCCTGAAAGTC
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAGTTGG
GTCCGGCAGCTCCA

#L34

CTCGAGTCTGGAGGTGGCCTGGTGCAGCCTGGAGGATCCCTCAAACTC
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAGTTGG
GTCCGGCAGCTCCA

#L50

CTCGAGTCAGGAGGTGGCCTGGTGCAGCCTGGAGGAGCCCTGAAACTC
TCCTGTGCAGCCTCAGGATTCGATTTnAGnAGATACTGGATGAGTTGG
GTCCGCAGCTCCAG

# FIG. 5-5

Subclass III (C)
#L10

CTCGAGTCTGGGGGAGGCTTAGTnCAGCCTGGAGGGTCCCGGAAACTC
TCCTGTGCAGCCTCTGGATTCACTTTnAGnAGTTTTGGAATGCACTGG
ATTCGTCAGGCTCC

#L08

CTCGAGTCTGGGGGAGGCTTAGTnnAGCCTGGAGGGTCCCGGAAACTC
TCCTGTGCAGCCTCTGGATTCACTTTnAGnAGCTTTGGAATGCACTGG
GTTACGTCAGGCTC

Subclass V (A)
#L38

CTCGAGTCAGGGGCTGAACTGGTGAGGCCTGGGCGTTCAGTnAAGATG
TCCTGCAAGGCTTCAGGCTATTCCTTCACCAGCTACTGGATGCACTGG
GTGAAACAGAGGCC

# FIG. 5-6

Miscellaneous
#47

CTCGAGTCAGGGGCTGAACTGGCAAAACCTGGGGCCTCAGTAAAGATG
TCCTGCAAGGCTTCTGGCTACACCTCTTCTTCCTTCTGGCTGCACTGG
ATAAAAGAAGGCCT

#L35

CTCGAGTCTGGACCTnAGCTGGTGAAGCCTGGGGTTCAGTTAAAATAT
CCTGCAAGGCTTCTGGTTACTCATTTTCTnTCTACTTTGTGAACTGGG
TGATGCAGAGCCAT

#L48

CTCGAGTCAGGGGCTGAACTGGTGAAGCCTGGGGTTCAGTAAGTTGTC
CTGAAGGCTTCTGGCTACACCTTCACCGGCTACTATATGTACTGGGTG
AAGCAGAGGCCTGG

# FIG. 6A-1

$V_H$ EXPRESSION VECTOR:

SHINE-DALGARNO MET

GGCCGCAAATTCTATTTCAAGGAGACAGTCATAATG
CGTTTAAGATAAAGTTCCTCTGTCAGTATTAC

LEADER SEQUENCE

AAATACCTATTGCCTACGGCAGCCGCT
TTTATGGATAACGGATGCCGTCGGCGA

LEADER SEQUENCE

GGATTGTTATTACTCGCTGCCCAACCAG
CCTAACAATAATGAGCGACGGGTTGGTC

# FIG. 6A-2

LINKER LINKER

NCOI $V_H$ BACKBONE XHOI SPEI

CCATGGCCCAGGTGAAACTGCTCGAGATTTCTAGACTAGT
GGTACCGGGTCCACTTTGACGAGCTCTAAAGATCTGATCA

STOP LINKER

TyrProTyrAspValProAspTyrAlaSer
TACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
ATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT

FIG. 6B-1

$V_L$ EXPRESSION VECTOR:

SHINE-DALGARNO

GGCCGCAAATTCTATTTCAAGGAGACAGTCATA
CGTTTAAGATAAAGTTCCTCTGTCAGTAT

MET LEADER SEQUENCE

ATGAAATACCTATTGCCTACGGCAGCCGCTGGA
TACTTTATGGATAACGGATGCCGTCGGCGACCT

LEADER SEQUENCE

TTGTTATTACTCGCTGCCCAACCAG
AACAATAATGAGCGACGGGTTGGTC

FIG. 6B-2

LINKER

NCOI $V_H$ BACKBONE

CCATGGCCCAGGTGAAACTG
GGTACCGGGTCCACTTTGAC

LINKER

XHOI SPEI STOP

CTCGAGAATTCTAGACTAGTTAATAG
GAGCTCTTAAGATCTGATCAATTATCAGCT

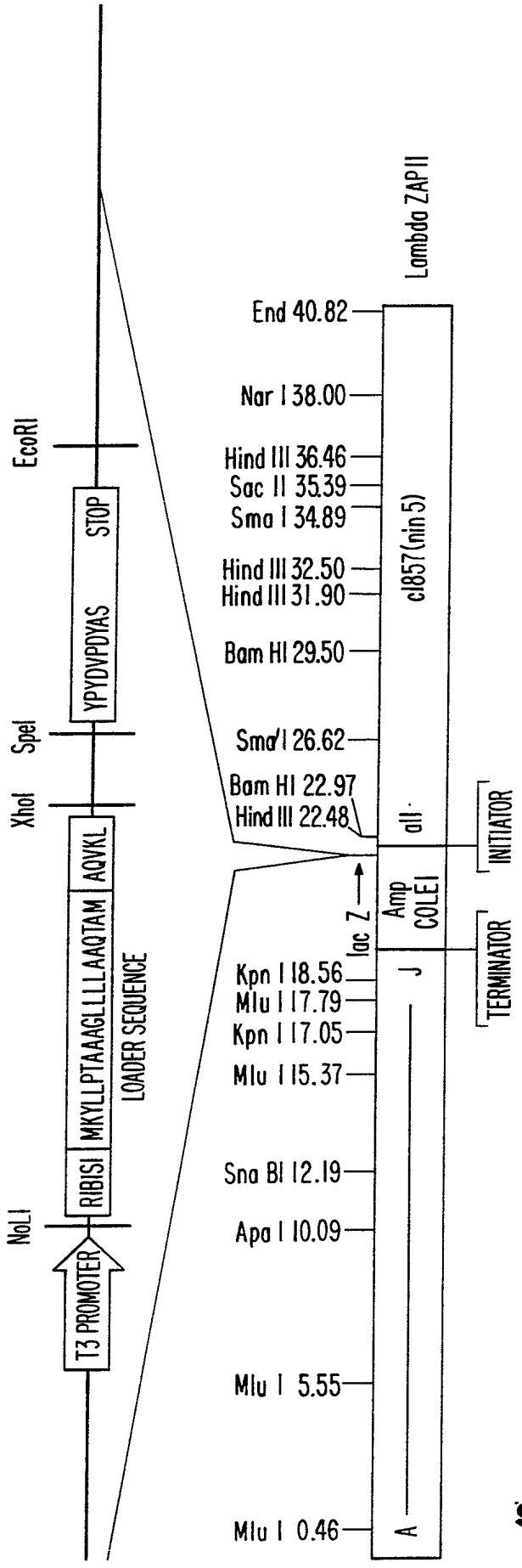



FIG. 7.

SUBSTITUTE SHEET

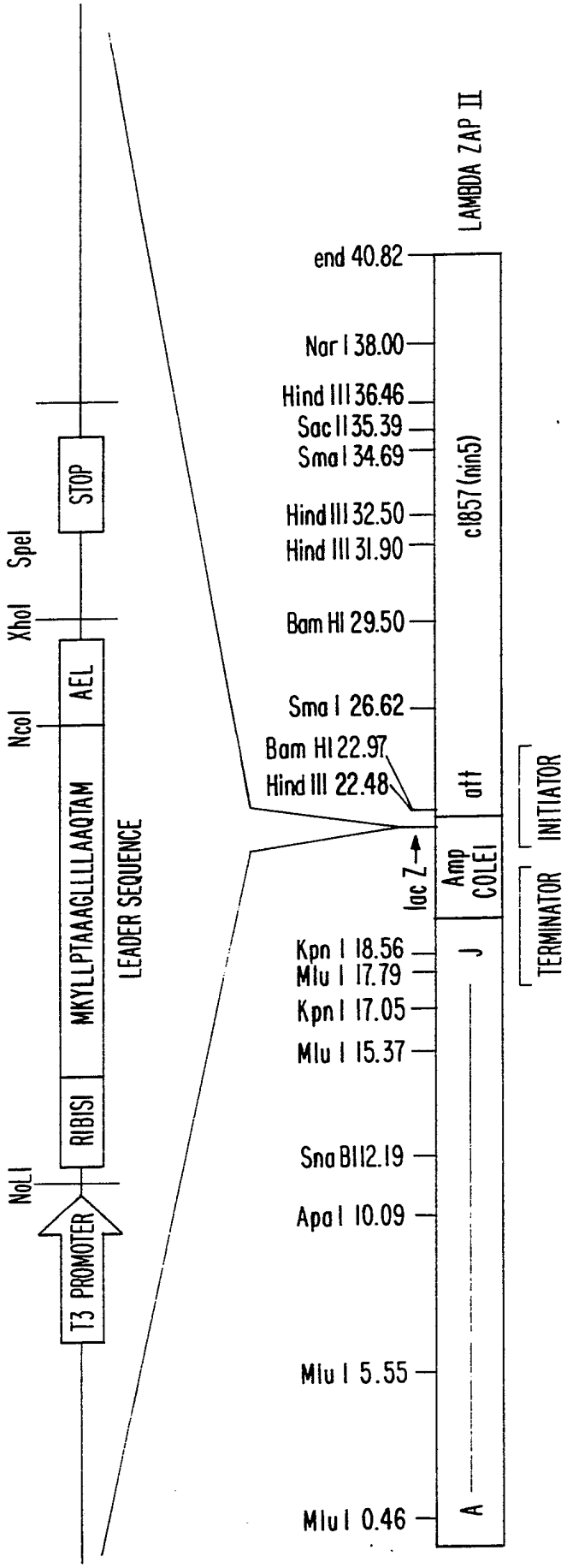

LAMBDA ZAP II
end 40.82
Nar I 38.00
Hind III 36.46
Sac II 35.39
Sma I 34.69
Hind III 32.50
Hind III 31.90
Bam HI 29.50
Sma I 26.62
Bam HI 22.97
Hind III 22.48
c1857 (nin5)
att
Amp
COLE1
lac Z
J
A
Kpn I 18.56
Mlu I 17.79
Kpn I 17.05
Mlu I 15.37
Sna BII 2.19
Apa I 10.09
Mlu I 5.55
Mlu I 0.46
INITIATOR
TERMINATOR
STOP
SpeI
XhoI
NcoI
NoLI
AEL
MKYLLPTAAAGLLLLAAQTAM
RIBISI
T3 PROMOTER
LEADER SEQUENCE

FIG. 8.

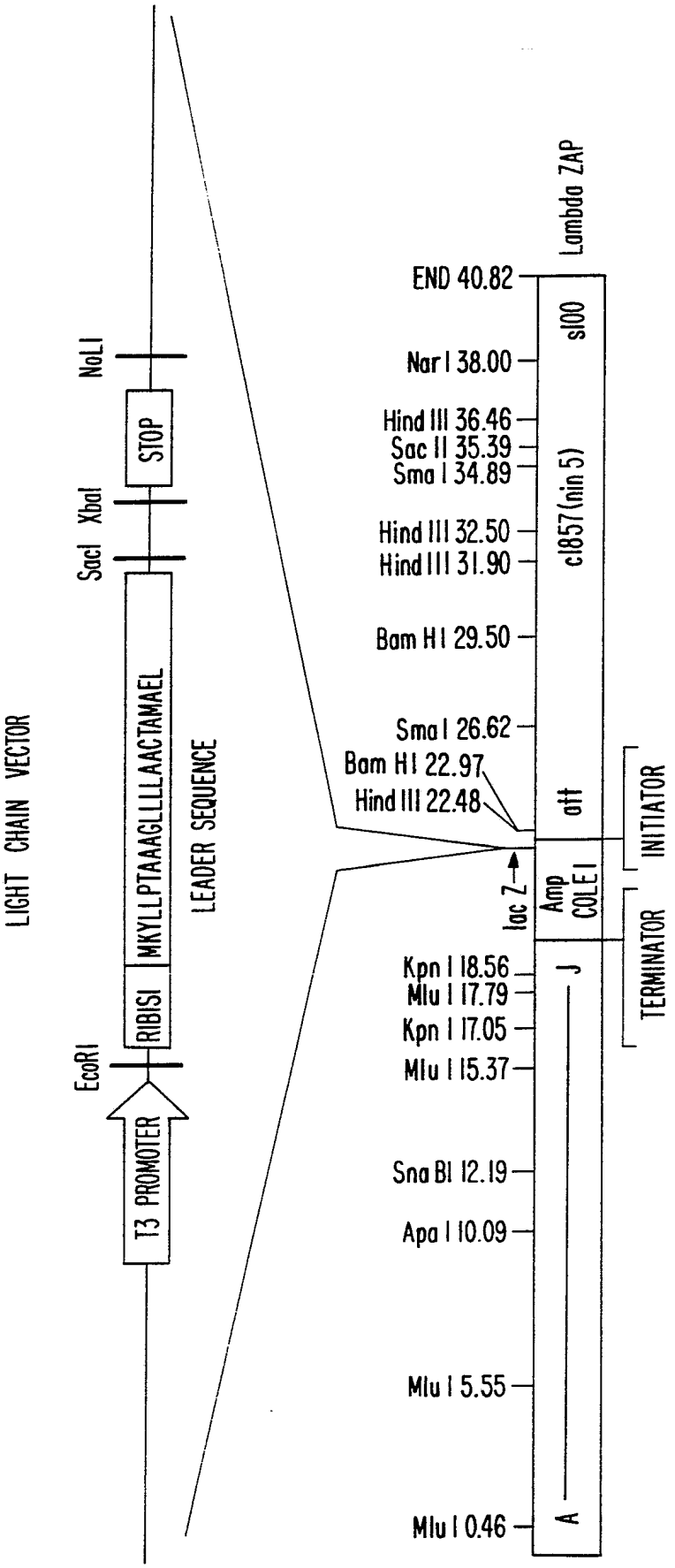
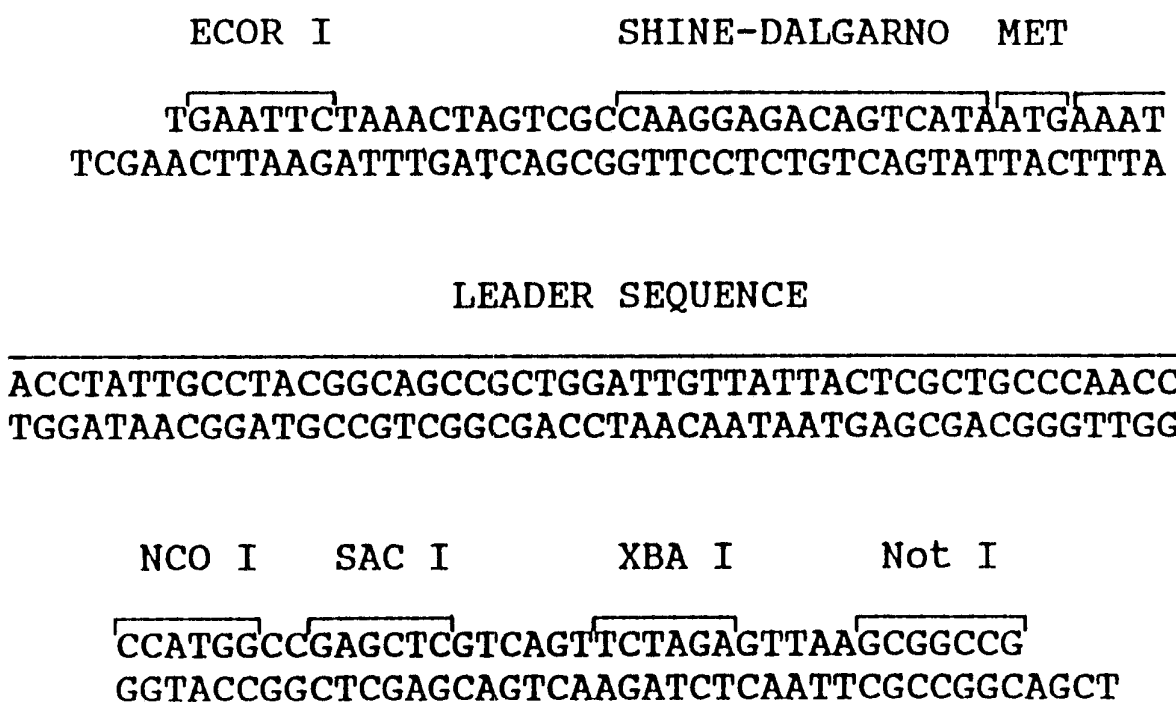

LIGHT CHAIN VECTOR
LEADER SEQUENCE
EcoRI
RIBISI
MKYLLPTAAGLLLLAACTAMAEL
SacI XbaI
NoLI
STOP
T3 PROMOTER
Lambda ZAP
END 40.82
Nar I 38.00
Hind III 36.46
Sac II 35.39
Sma I 34.89
Hind III 32.50
Hind III 31.90
Bam HI 29.50
Sma I 26.62
Bam HI 22.97
Hind III 22.48
sI00
c1857(nin 5)
att
INITIATOR
lac Z
Amp
COLE1
TERMINATOR
J
A
Kpn I 18.56
Mlu I 17.79
Kpn I 17.05
Mlu I 15.37
Sna BI 12.19
Apa I 10.09
Mlu I 5.55
Mlu I 0.46

FIG. 9.

# FIG. 10



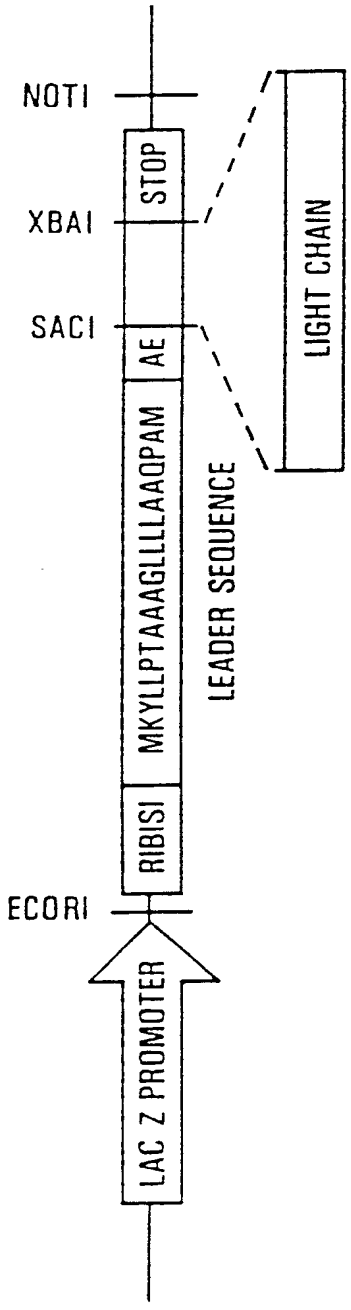



FIG. 11-1

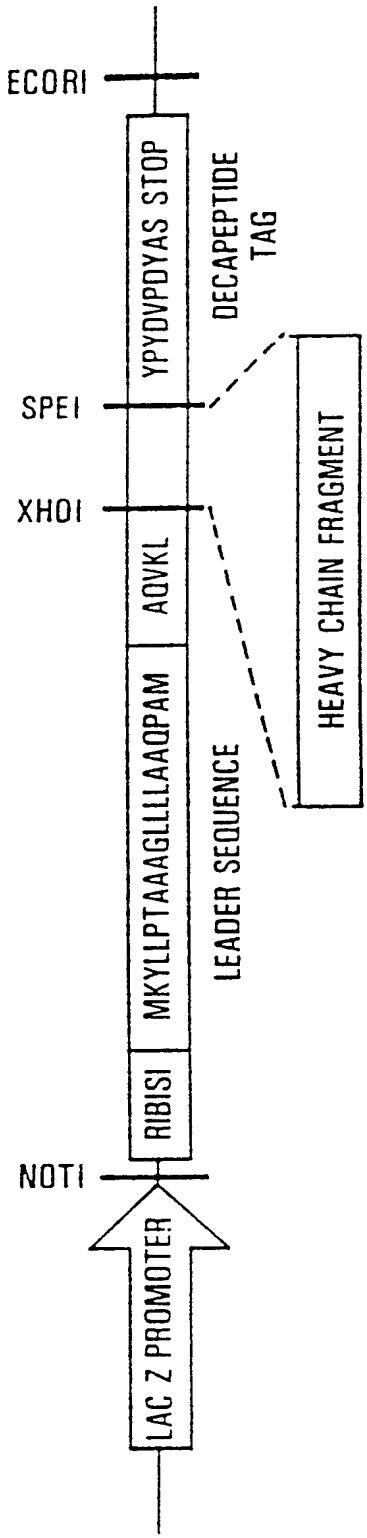



FIG. 11-2

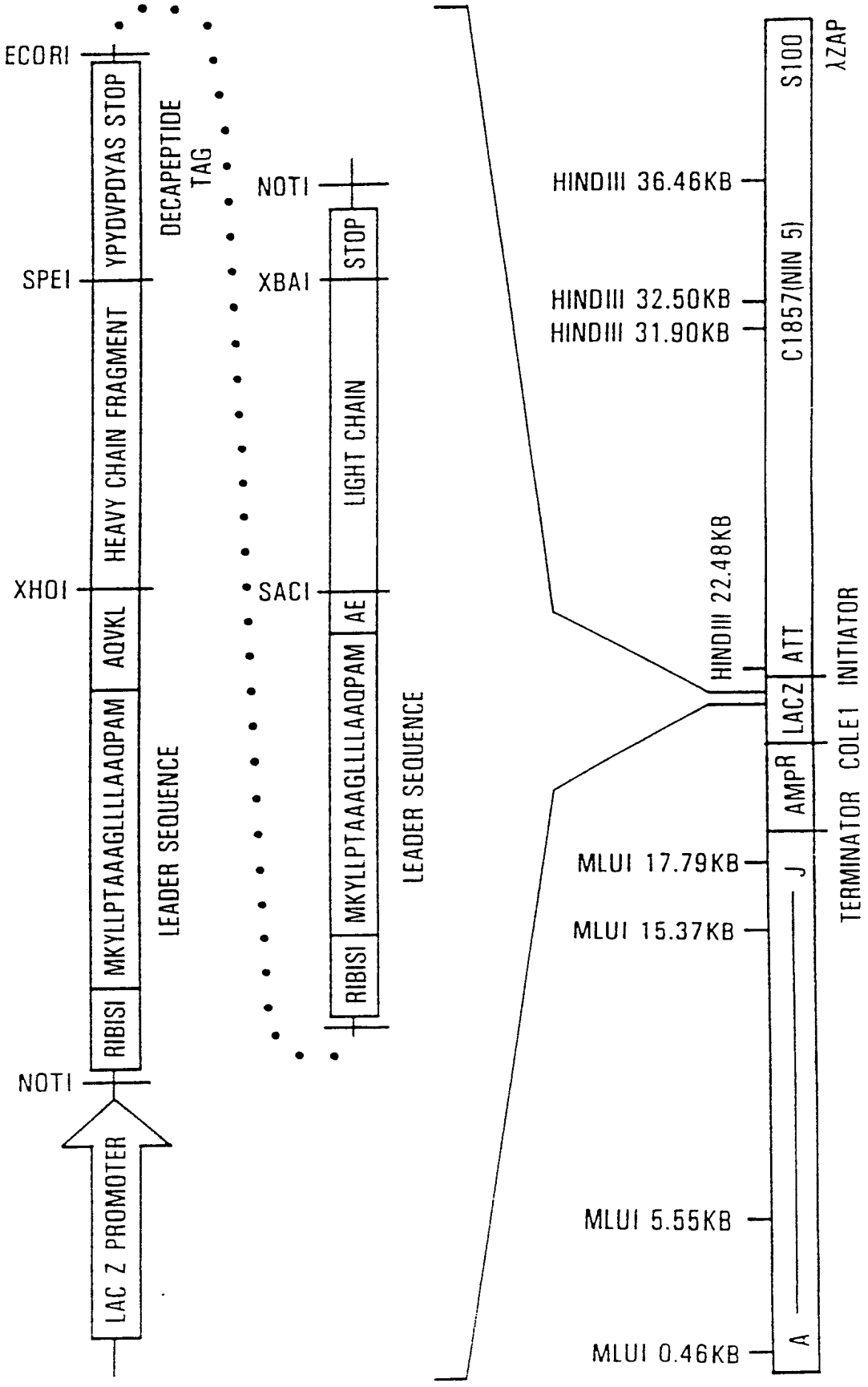

ECORI
SPEI
XHOI
NOTI
LAC Z PROMOTER
RIBISI
MKYLLPTAAAGLLLLAAQPAM
AQVKL
HEAVY CHAIN FRAGMENT
YPYDVPDYAS STOP
LEADER SEQUENCE
DECAPEPTIDE TAG
NOTI
XBAI
SACI
RIBISI
MKYLLPTAAAGLLLLAAQPAM
AE
LIGHT CHAIN
STOP
LEADER SEQUENCE
HINDIII 36.46KB
HINDIII 32.50KB
HINDIII 31.90KB
HINDIII 22.48KB
MLUI 17.79KB
MLUI 15.37KB
MLUI 5.55KB
MLUI 0.46KB
A
J
AMPR
LACZ
ATT
C1857(NIN 5)
S100
λZAP
TERMINATOR COLE1 INITIATOR

FIG. 11-3

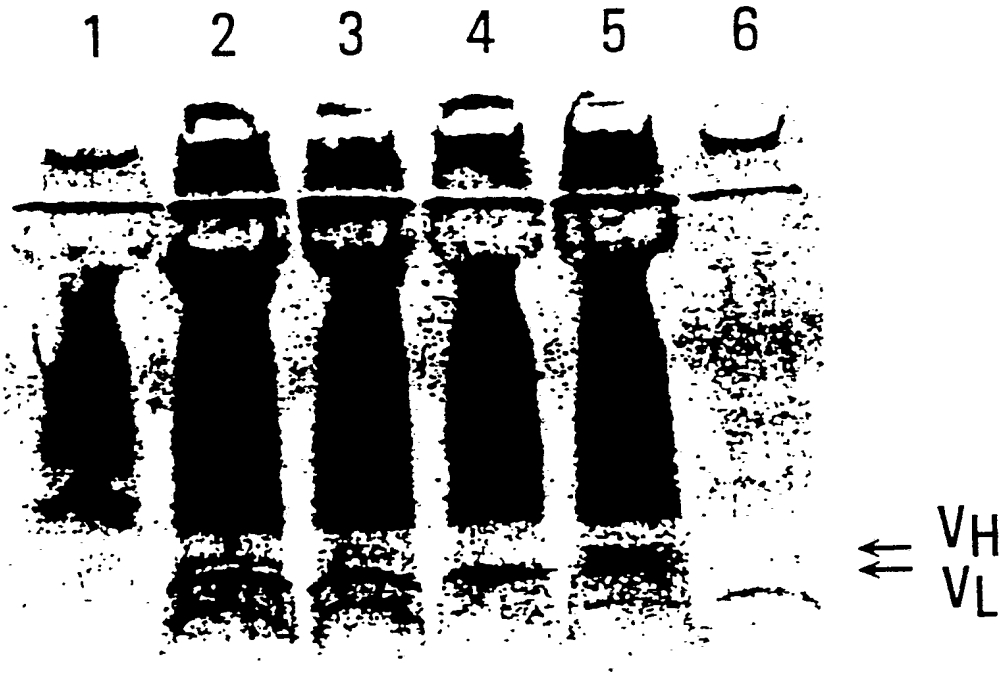



FIG. 12

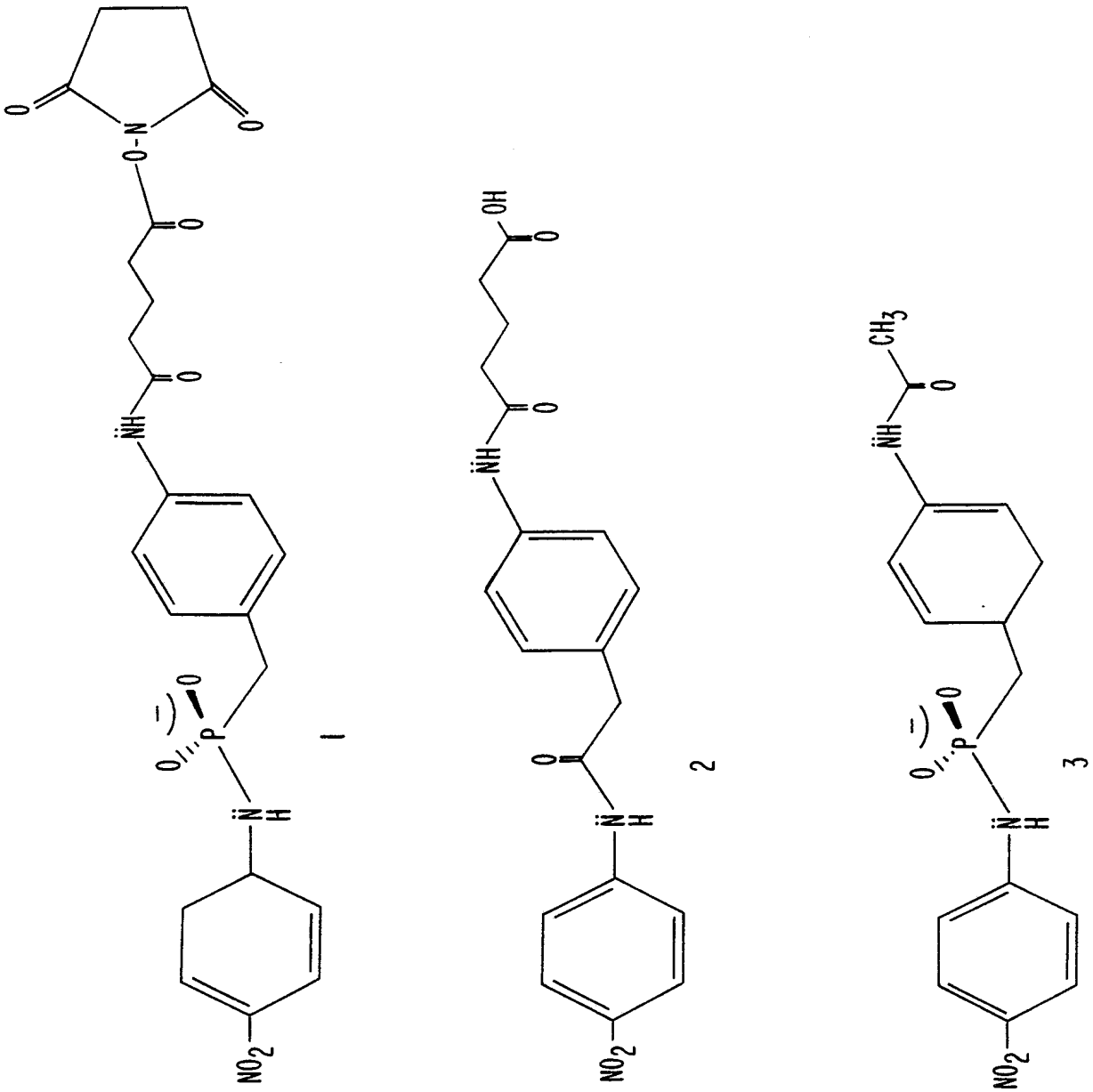


FIG. 13.

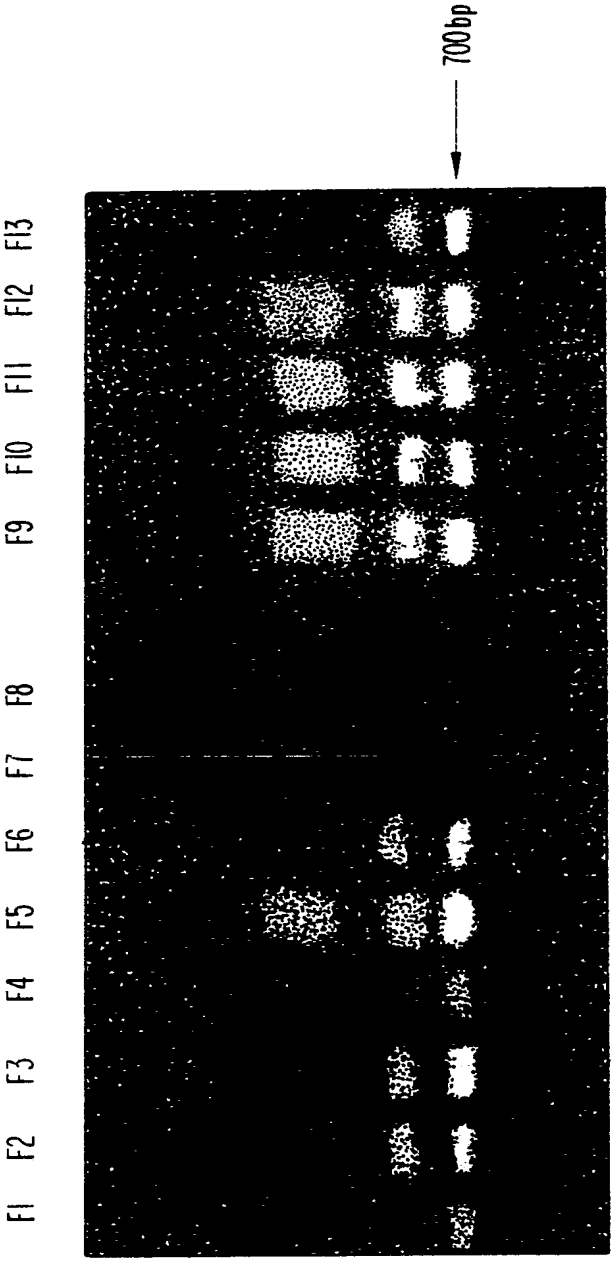



FIG. 14.

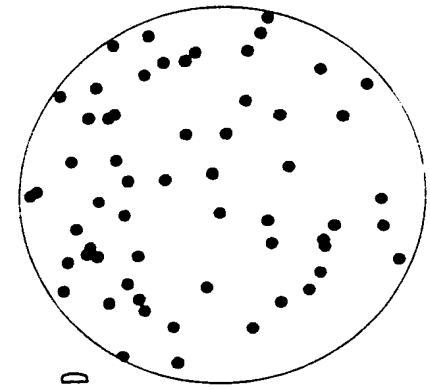



FIG. 15d.

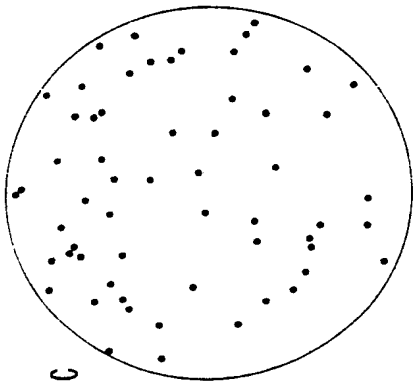



FIG. 15c.

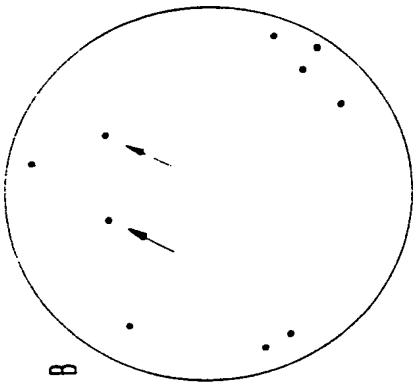



FIG. 15b

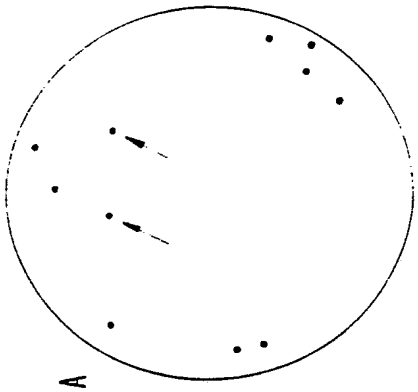



FIG. 15a.

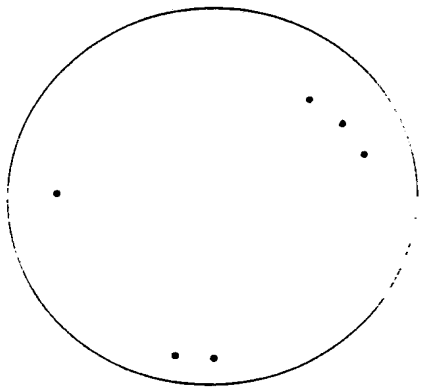



FIG. 15h.

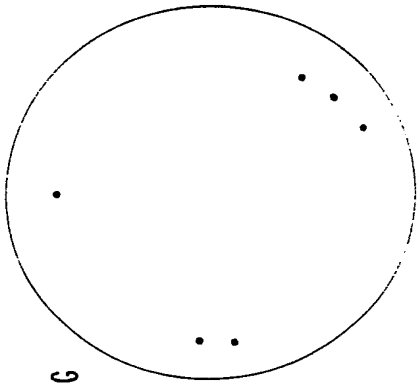



FIG. 15g.

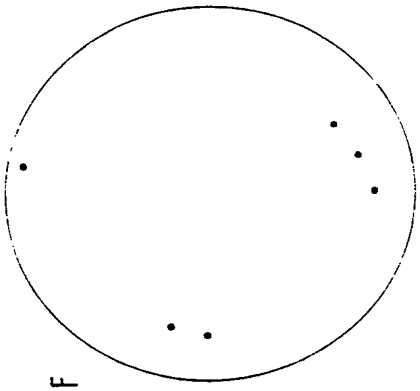



FIG. 15f.

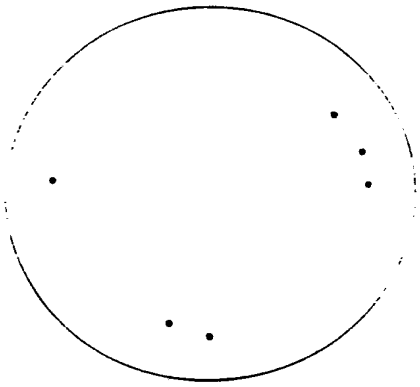
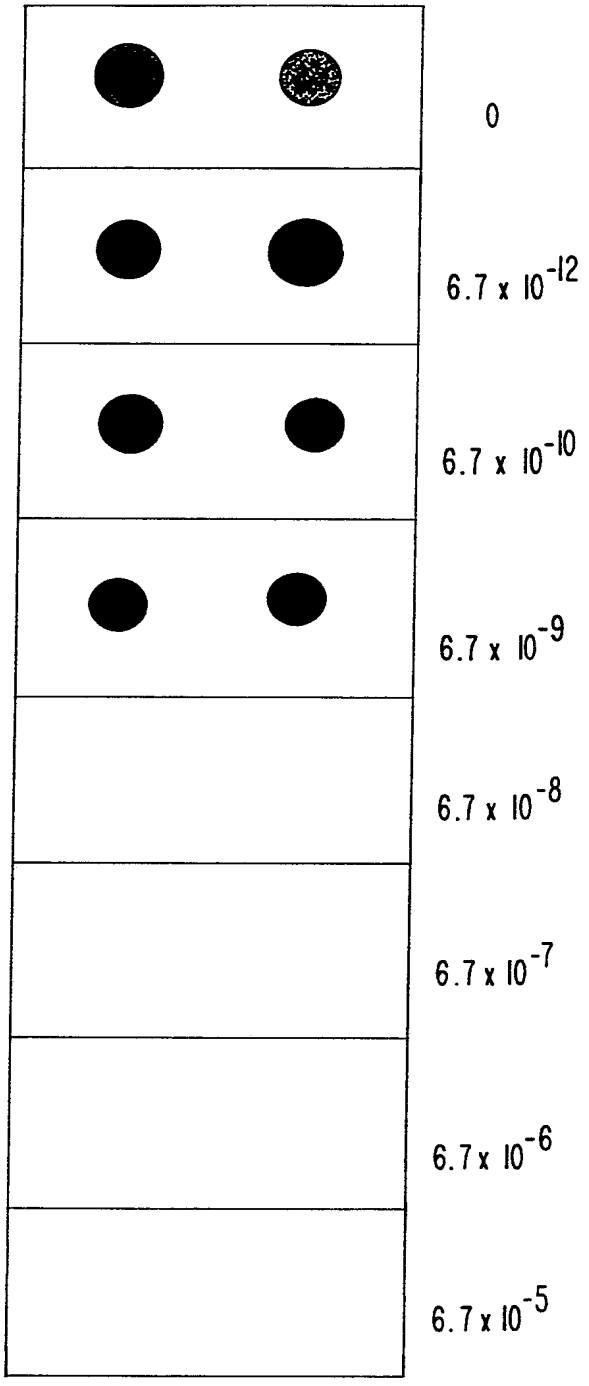



FIG. 15e.

FIG. 16.



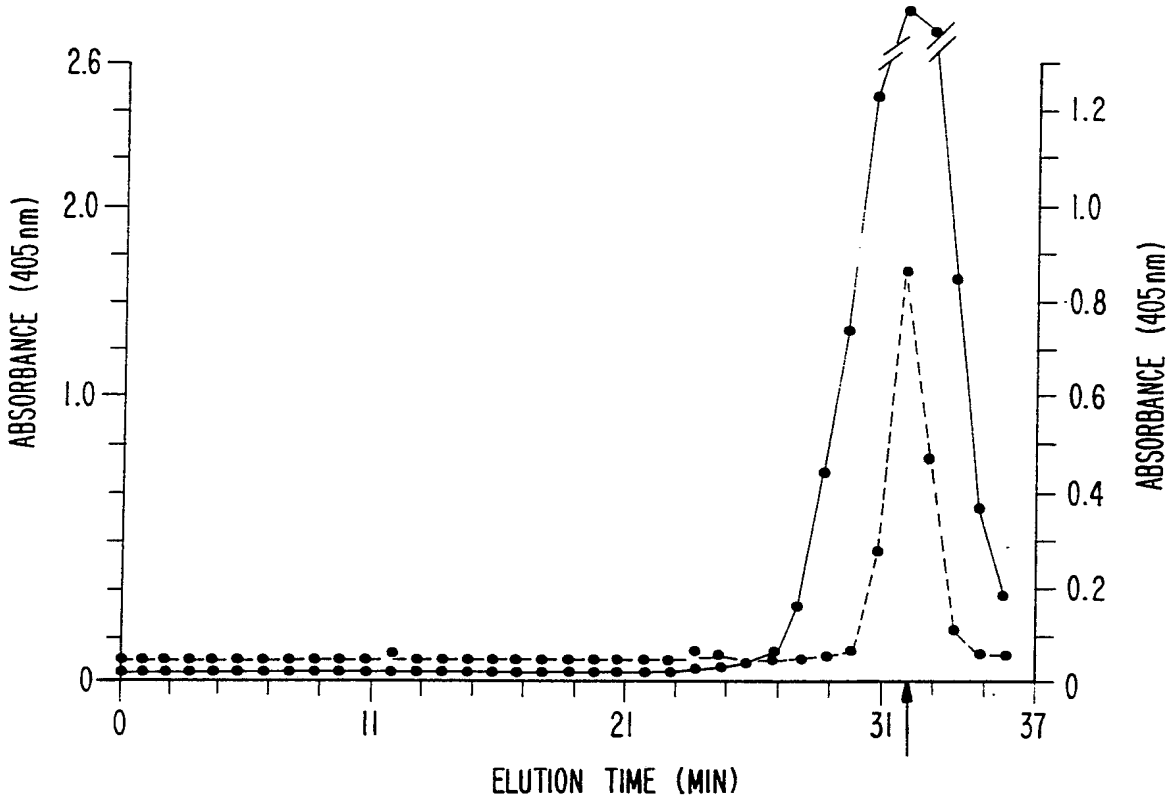

2.6
2.0
1.0
0
ABSORBANCE (405 nm)
1.2
1.0
0.8
0.6
0.4
0.2
0
ABSORBANCE (405 nm)
0
11
21
31
37
ELUTION TIME (MIN)

FIG. 17.

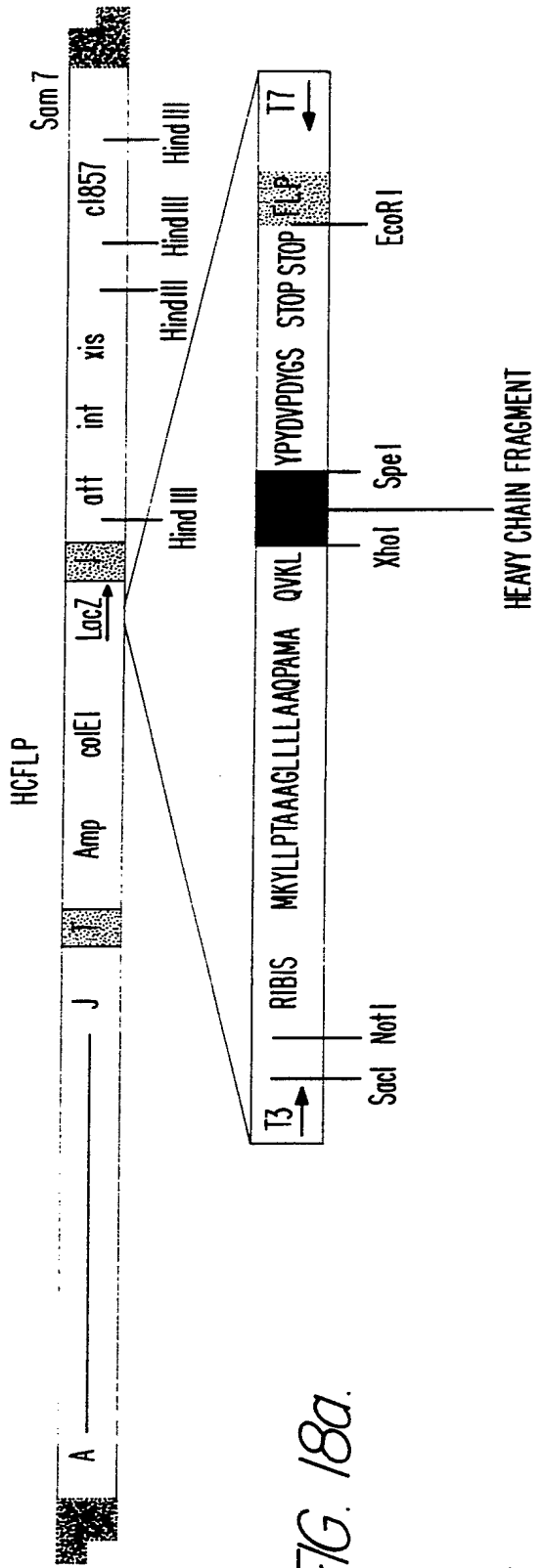



FIG. 18a.

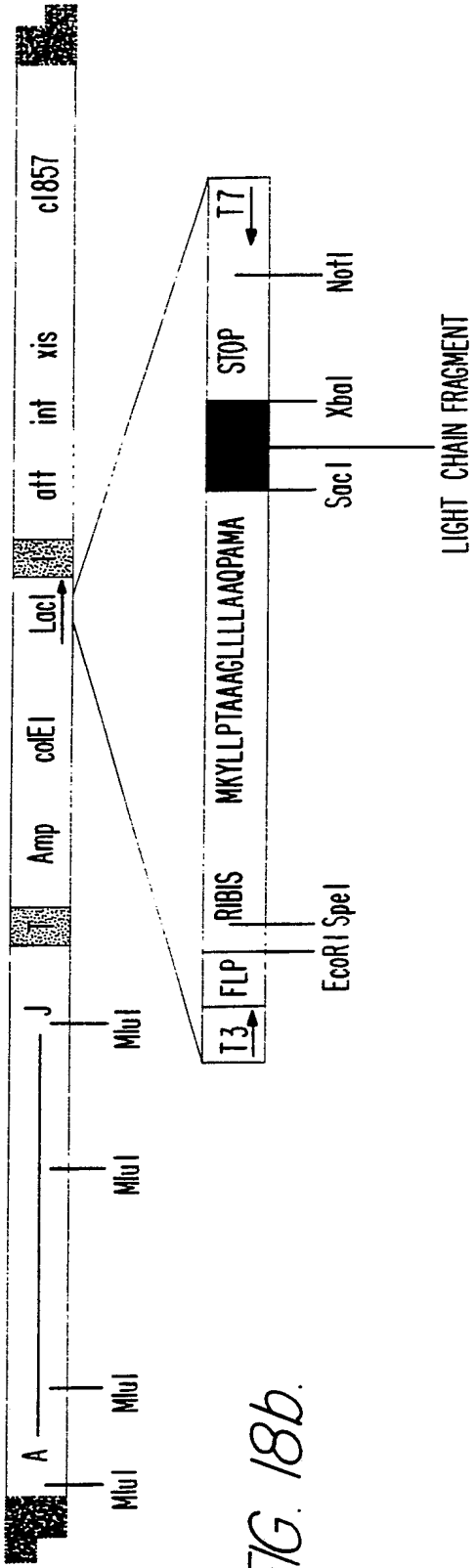



FIG. 18b.

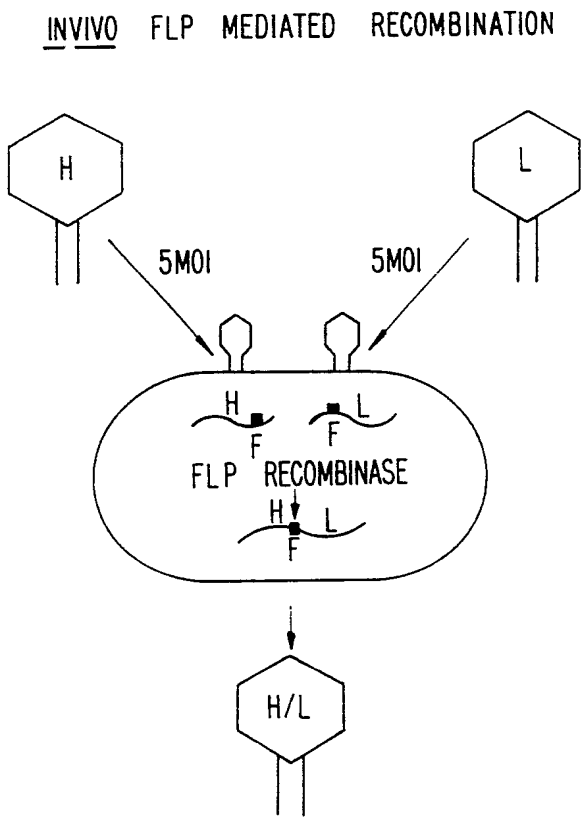

INVIVO FLP MEDIATED RECOMBINATION
H
L
5MOI
5MOI
H
F
L
F
FLP RECOMBINASE
H
L
F
H/L

FIG. 19.

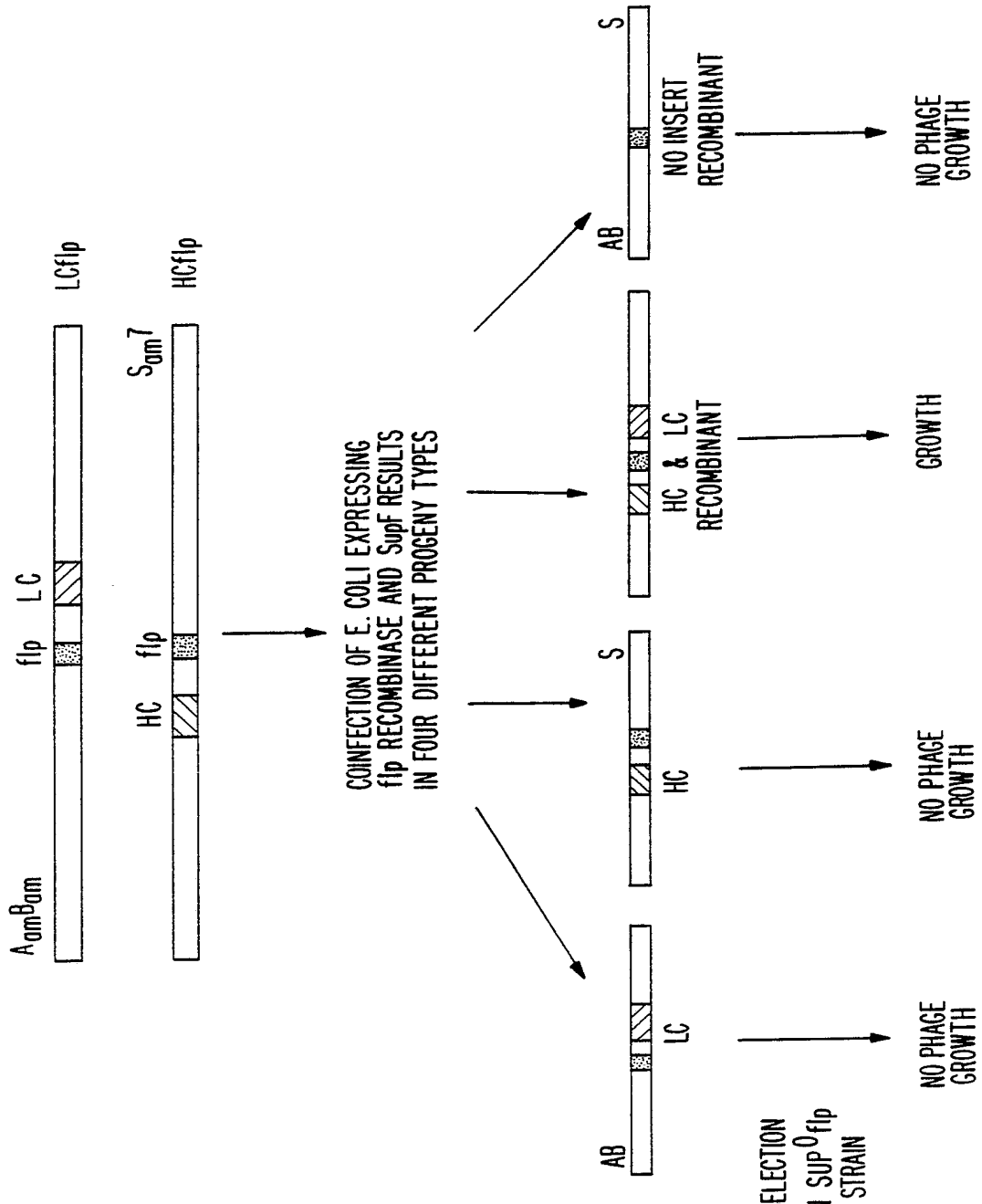

LCflp
HCflp
flp LC
HC flp
$A_{am}B_{am}$
$S_{am}7$
COINFECTION OF E. COLI EXPRESSING
flp RECOMBINASE AND SupF RESULTS
IN FOUR DIFFERENT PROGENY TYPES
AB
S
NO INSERT
RECOMBINANT
NO PHAGE
GROWTH
HC & LC
RECOMBINANT
GROWTH
HC
NO PHAGE
GROWTH
LC
NO PHAGE
GROWTH
SELECTION
ON $SUP^0$ flp
STRAIN

FIG. 20.

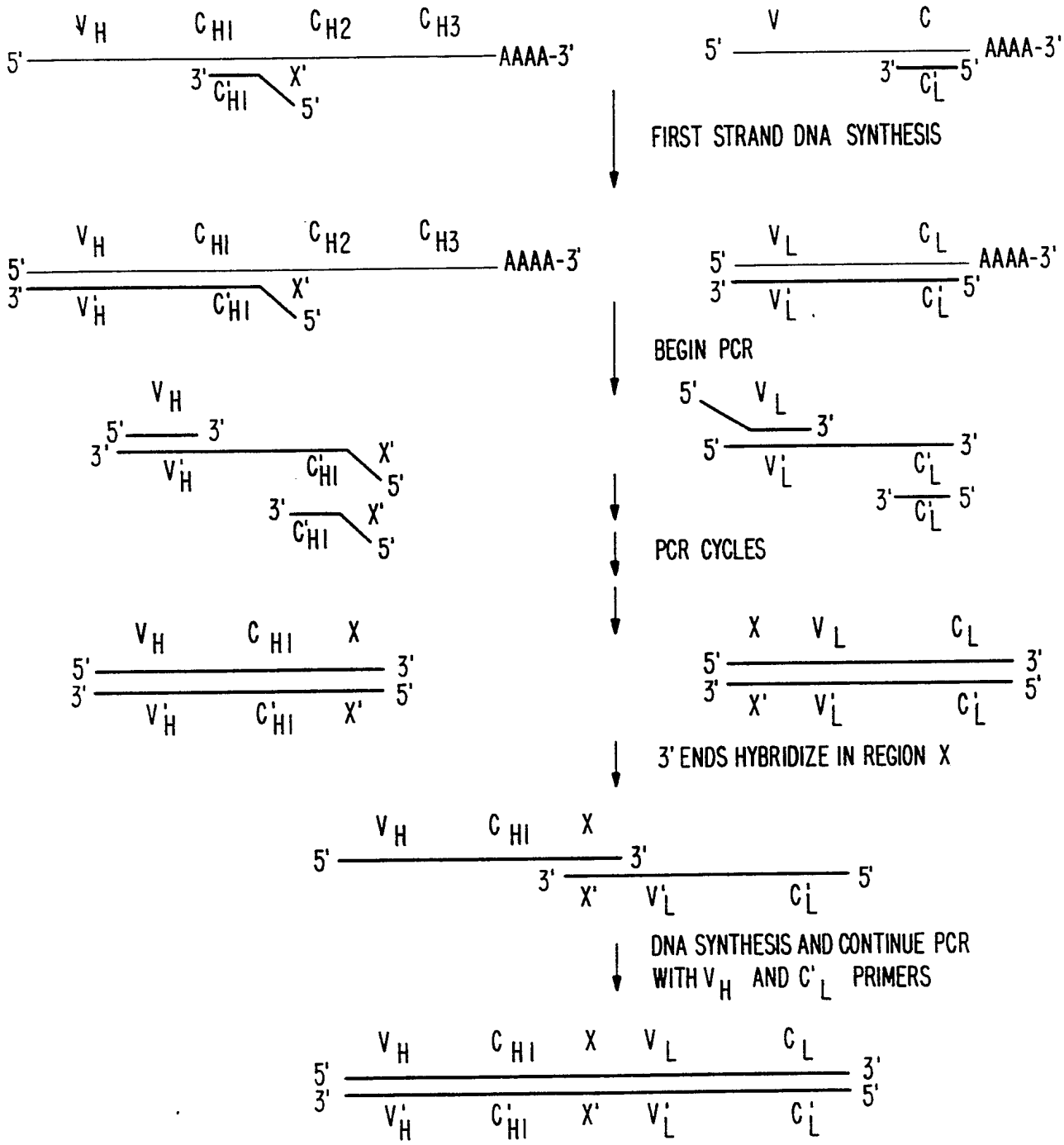

FIRST STRAND DNA SYNTHESIS
BEGIN PCR
PCR CYCLES
3' ENDS HYBRIDIZE IN REGION X
DNA SYNTHESIS AND CONTINUE PCR
WITH V_H AND C'_L PRIMERS

FIG. 21.

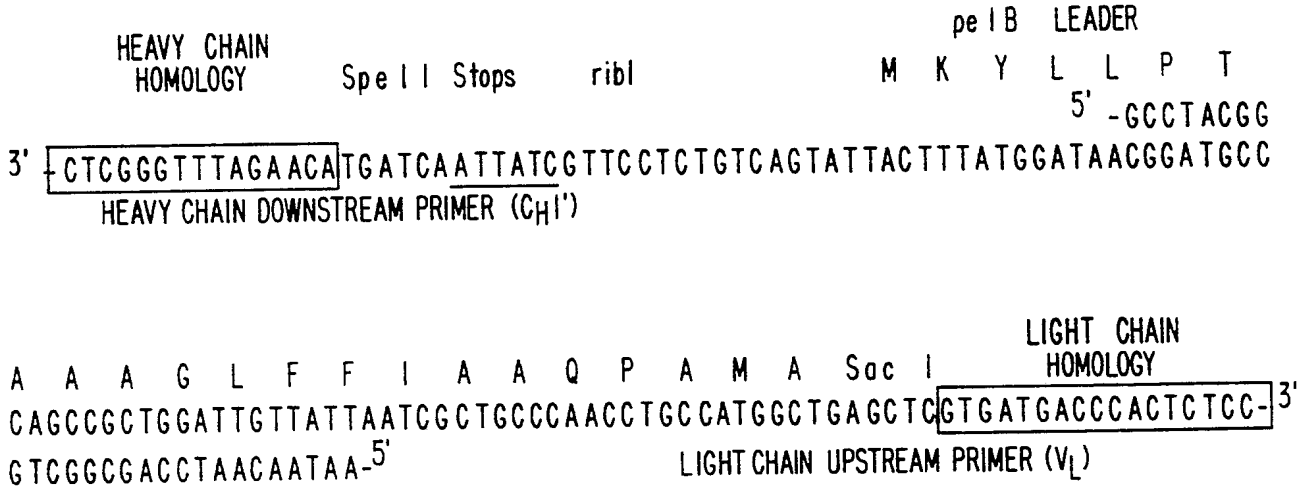

HEAVY CHAIN HOMOLOGY
Spe I I Stops
ribl
pe I B LEADER
M K Y L L P T
5' -GCCTACGG
3' -CTCGGGTTTAGAACATGATCAATTATCGTTCCTCTGTCAGTATTACTTTATGGATAACGGATGCC
HEAVY CHAIN DOWNSTREAM PRIMER (CH1')
A A A G L F F I A A Q P A M A Sac I
LIGHT CHAIN HOMOLOGY
CAGCCGCTGGATTGTTATTAATCGCTGCCCAACCTGCCATGGCTGAGCTCGTGATGACCCACTCTCC- 3'
GTCGGCGACCTAACAATAA-5'
LIGHT CHAIN UPSTREAM PRIMER (VL)

FIG. 22.

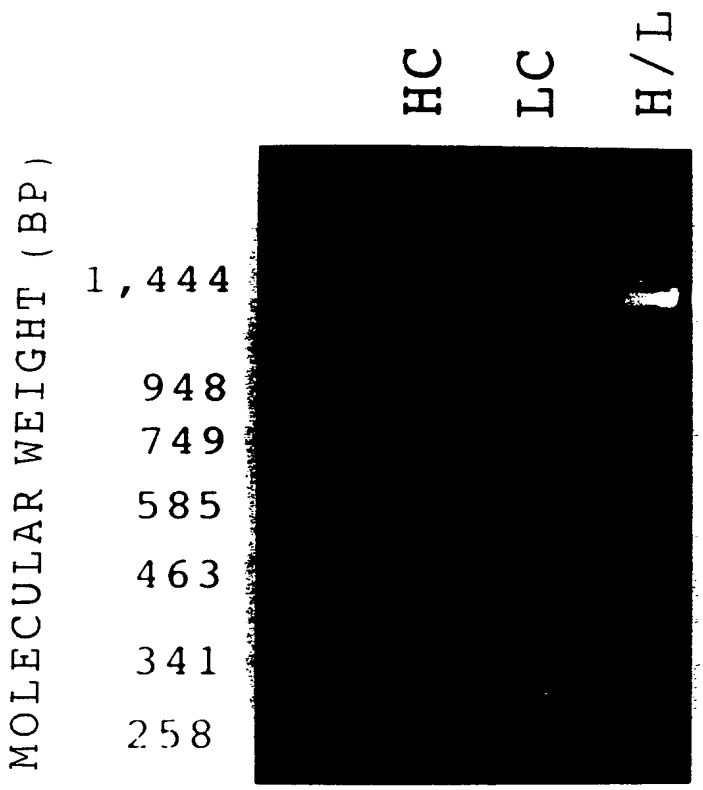



FIG. 23.

MODIFIED $V_H$ EXPRESSION VECTOR:

Not I RIBOSOME BINDING SITE

5' GAGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCATA
3' CGCCGGCGTTTAAGATTAAAGTTCCTCTGTCAGTAT

Pel B LEADER

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA

FIG. 24a.

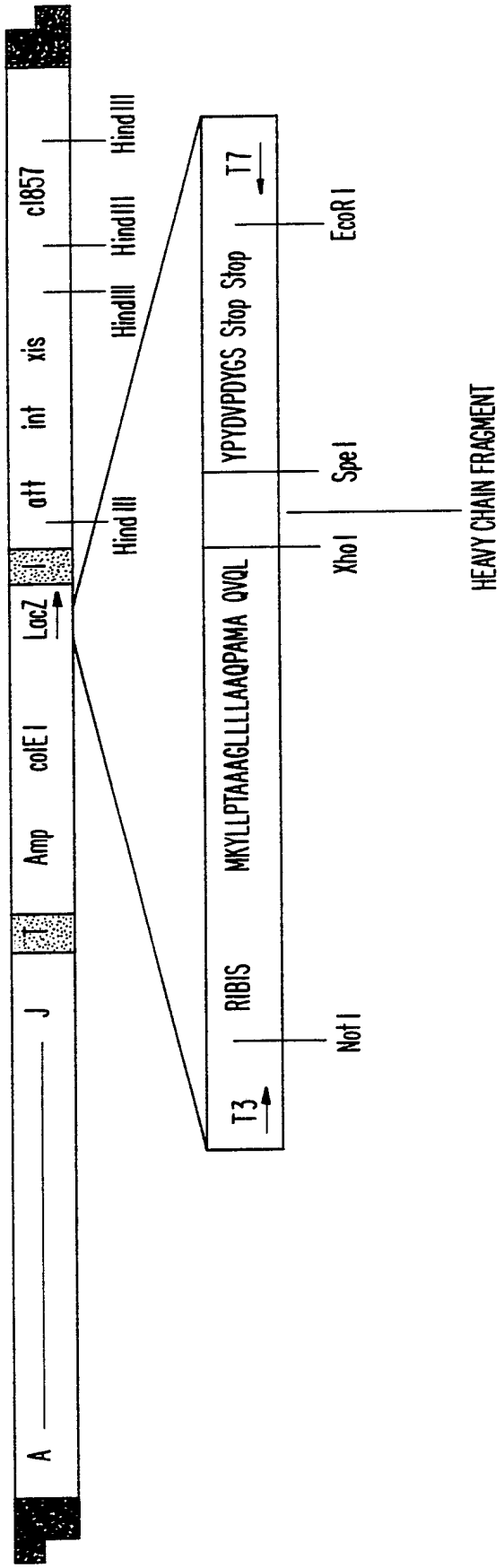
Nco I Xho I Xba I Spe I

AlaGlnProAlaMetAlaGlnValGlnLeuLeuGlu Thr
GCCCAACCAGCCATGGCCCAGGTGCAGCTGCTCGAGATTTCTAGACT
CGGGTTGGTCGGTACCGGGTCCACGTCGACGAGCTCTAAAGATCTGA

EcoRI

SerTyrProTyrAspValProAspTyrGlySerStop
AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT

FIG. 24b.


A
J
T1
Amp
colE1
LacZ
I
att
int
xis
cI857
HindIII
HindIII
HindIII
HindIII
RIBIS
MKYLLPTAAAGLLLLAAQPAMA QVQL
YPYDVPDYGS Stop Stop
T3
T7
NotI
XhoI
SpeI
EcoRI
HEAVY CHAIN FRAGMENT

$V_H$ EXPRESSION VECTOR:

Not I RIBOSOME BINDING SITE

5' GGCCGCAAATTCTATTTCAAGGAGACAGTCATA
CGTTTAAGATAAAGTTCCTCTGTCAGTAT

Pel B LEADER

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA

NcoI XhoI XbaI SpeI

$V_H$ BACKBONE

AlaGlnProAlaMetAlaGlnValLysLeuLeuGlu Thr
GCCCAACCAGCCATGGCCCAGGTGAAACTGCTCGAGTTTCTAGACT
CGGGTTGGTCGGTACCGGGTCCACTTTGACGAGCTCTAAAGATCTGA

EcoRI

SerTyrProTyrAspValProAspTyrGlySerStop
AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT

FIG. 25a.

$V_L$ EXPRESSION VECTOR:

EcoRI RIBOSOME BINDING SITE

5' TGAATTCTAAACTAGTCGCCAAGGAGACAGTCATA
3' TCGAACTTAAGATTTGATCAGCGGTTCCTCTGTCAGTAT

PeI B LEADER

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeu
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTC
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAG

FIG. 25b.

NcoI SacI

AlaAlaGlnProAlaMetAlaGluLeu
GCTGCCCAACCAGCCATGGCCGAGCTC
CGACGGGTTGGTCGGTACCGGCTCGAG

XbaI
Stop Stop

GTCAGTTCTAGAGTTAAGCGGCCG
CAGTCAAGATCTCAATTCGCCGGCAGCT

FIG. 26.

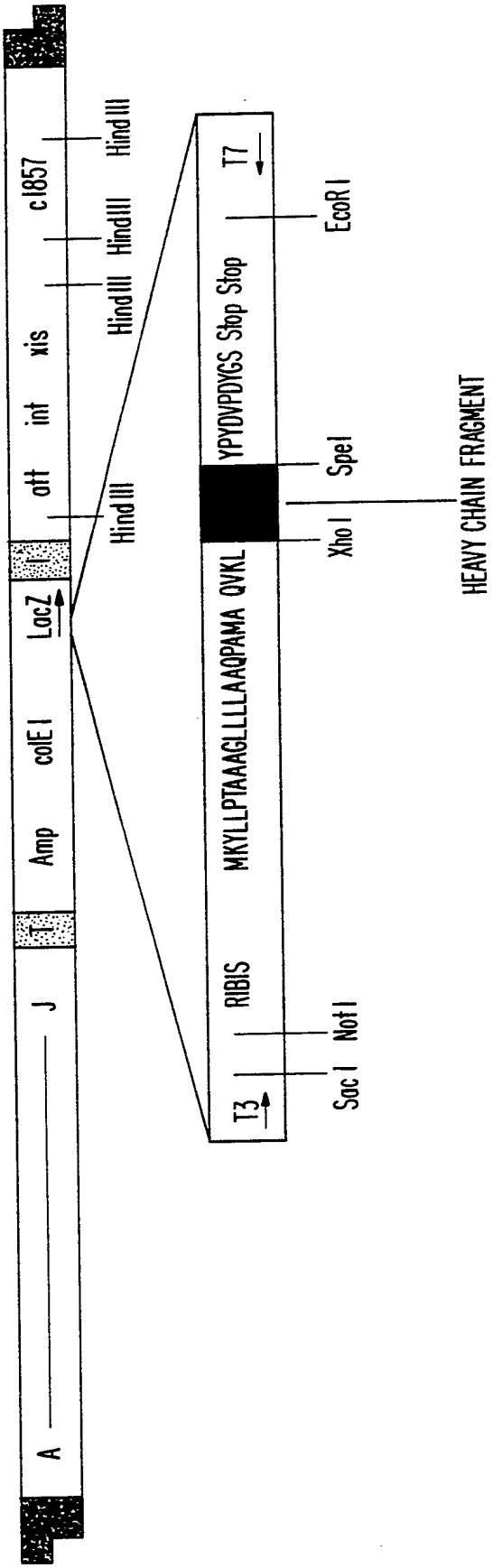
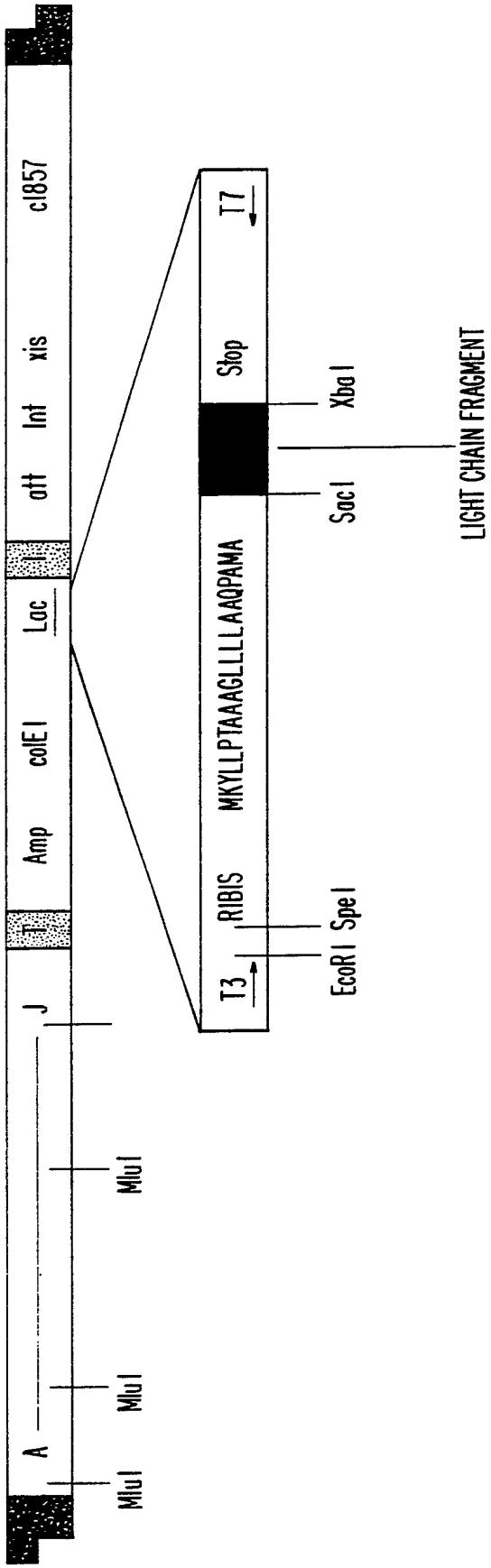

A
J
T1
Amp
colE1
LacZ
I
ott
int
xis
c1857
HindIII
HindIII
HindIII
HindIII
T3
T7
RIBIS
MKYLLPTAAAGLLLLAAQPAMA QVKL
YPYDVPDYGS Stop Stop
Sac I
Not I
Xho I
Spe I
EcoR I
HEAVY CHAIN FRAGMENT

FIG. 27.


c1857
att
Int
xis
Loc
colE1
Amp
J
A
MluI
MluI
MluI
T7
Stop
XbaI
SacI
LIGHT CHAIN FRAGMENT
MKYLLPTAAAGLLLLAAQPAMA
RIBIS
SpeI
EcoRI
T3

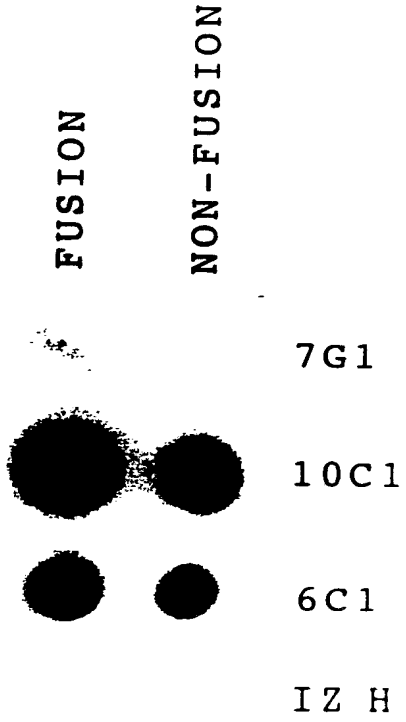



FIG. 28.

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/02910

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

US: 435/172.1,172.2,172.3,69.6,69.7,91; 536/27; 935/22,23

IPC(5): C12N 15/00; C12P 19/34, 21/06; C07H 21/00

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

| Classification System | Classification Symbols |
|---|---|
| US | 435/172.3, 91 69.6 69.7,172.1,172.2; 935/22,23; 536/57 |

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

APS

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

| Category * | Citation of Document, [11] with indication, where appropriate, of the relevant passages [12] | Relevant to Claim No. [13] |
|---|---|---|
| Y | Proceedings of the National Academy of Sciences. Volume 82. issued November 1985, J.F. Senecoff et al "the FLP recombinase of the yeast 2-um plasmid: characterization of its recombination site". pages 7270-7274. See entire article. | 1-10,13 14-40 59-64 |
| Y | US.A. 4,683.195 (Mullis et al) 28 July 1987. See entire document. | 62-64 |
| Y | US.A. 4.642.334 (Moore et al) 10 February 1987. See entire document. | 1-14,19-58, |
| Y | US.A. 4,816,397 (Boss et al) 28 March 1989 See entire document. | 1-14,19-58, 65-76 |

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

| Date of the Actual Completion of the International Search | Date of Mailing of this International Search Report |
|---|---|
| 24 July 1991 | 16 AUG 1991 |
| International Searching Authority | Signature of Authorized Officer |
| ISA/US | Suzanne Ziska |